| | |
|---|---|
| **Titre:** Title: | Fault Prognostics Using Logical Analysis of Data and Non-Parametric Reliability Estimation Methods |
| **Auteur:** Author: | Ahmed Ragab Anwar Ragab |
| **Date:** | 2014 |
| **Type:** | Mémoire ou thèse / Dissertation or Thesis |
| **Référence:** Citation: | Ragab, A. R. A. (2014). Fault Prognostics Using Logical Analysis of Data and Non-Parametric Reliability Estimation Methods [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/1658/ |



| | |
|---|---|
| **URL de PolyPublie:** PolyPublie URL: | https://publications.polymtl.ca/1658/ |
| **Directeurs de recherche:** Advisors: | Soumaya Yacout, & Mohamed-Salah Ouali |
| **Programme:** Program: | Génie industriel |

UNIVERSITÉ DE MONTRÉAL

# FAULT PROGNOSTICS USING LOGICAL ANALYSIS OF DATA AND NON-PARAMETRIC RELIABILITY ESTIMATION METHODS

AHMED RAGAB ANWAR RAGAB

DÉPARTEMENT DE MATHÉMATIQUES ET DE GENIE INDUSTRIEL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR

(GÉNIE INDUSTRIEL)

DECEMBRE 2014

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

FAULT PROGNOSTICS USING LOGICAL ANALYSIS OF DATA AND NON-PARAMETRIC RELIABILITY ESTIMATION METHODS

présentée par : RAGAB Ahmed Ragab Anwar

en vue de l'obtention du diplôme de : Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de :

M. ADJENGUE Luc-Désiré, Ph. D., président

Mme YACOUT Soumaya, D. Sc., membre et directrice de recherche

M. OUALI Mohamed-Salah, Doctorat, membre et codirecteur de recherche

M. AGARD Bruno, Doctorat, membre

M. KUMRAL Mustafa, Ph. D., membre

# DEDICATE

*I dedicate this thesis to my parents, my beloved wife (Yasmine), my son (Omar), and my daughter (Sarah) whom I am indebted to for their sacrifices, encouragement, and endless and unwavering love.*

# ACKNOWLEDGEMENTS

There are many people I wish to thank for their help on this thesis. First, I would like to express my profound gratitude to my research director Dr. Soumaya Yacout whom I have had the privilege of being her student and whom I am heavily indebted to. Without her, this thesis could not have been found. I am even more fortunate that she generously gave me more of her valuable time to complete this work. I am indebted to her initial vision and long-term support for the research interest upon which this thesis is based. I appreciate her patience and gentle guidance during the whole process of this research. Her words of wisdom provided me with enough confidence in my skills and encouraged me to undertake this work. She shared with me her knowledge that is valuable as gold.

My profound gratitude goes as well to my co-director of research Dr. Mohamed-Salah Ouali whom I received wisdom and knowledge. He shared with me his experience and knowledge throughout the work. I appreciate his dedication and insightful comments. I have relied upon his insights, encouragement, and experience to formulate my thinking with regard to the technical and theoretical aspects in this thesis. I am deeply grateful to his remarkable support. I appreciate his patience, suggestions for improvements, and willingness to test algorithms and point out errors in the programs' codes. I am grateful to him for giving me the chance to prove myself.

I owe a special note of regards to Dr. Luc Adjengue, Dr. Bruno Agard, and Dr. Mustafa Kumral for their constructive role as members of my jury.

My gratitude and appreciation go as well to Dr. Hany Osman for providing me continuously with the updated versions of the software *cbmLAD*.

I would like to extend my thanks and appreciation to all those colleagues from our department for their support and advices to enhance the compilation of this thesis, in particular Xavier de Carné de Carnavalet, Mohamed Ossama, and Wissem Maazoun.

I would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for their fund to accomplish this doctoral program. My thanks go as well to Prüftechnik Canada for supporting us with the vibration data employed in this research.

# RÉSUMÉ

Estimer la durée de vie utile restante (RUL) d’un système qui fonctionne suivant différentes conditions de fonctionnement représente un grand défi pour les chercheurs en maintenance conditionnelle (CBM). En effet, il est difficile de comprendre la relation entre les variables qui représentent ces conditions de fonctionnement et la RUL dans beaucoup de cas en pratique à cause du degré élevé de corrélation entre ces variables et leur dépendance dans le temps. Il est également difficile, voire impossible, pour des experts d’acquérir et accumuler un savoir à propos de systèmes complexes, où l'échec de l'ensemble du système est vu comme le résultat de l'interaction et de la concurrence entre plusieurs modes de défaillance.

Cette thèse présente des méthodologies pour le pronostic en CBM basé sur l'apprentissage automatique, et une approche de découverte de connaissances appelée Logical Analysis of Data (LAD). Les méthodologies proposées se composent de plusieurs implémentations de la LAD combinées avec des méthodes non paramétriques d'estimation de fiabilité. L'objectif de ces méthodologies est de prédire la RUL du système surveillé tout en tenant compte de l'analyse des modes de défaillance uniques ou multiples. Deux d’entre elles considèrent un mode de défaillance unique et une autre considère de multiples modes de défaillance. Les deux méthodologies pour le pronostic avec mode unique diffèrent dans la manière de manipuler les données.

Les méthodologies de pronostique dans cette recherche doctorale ont été testées et validées sur la base d'un ensemble de tests bien connus. Dans ces tests, les méthodologies ont été comparées à des techniques de pronostic connues; le modèle à risques proportionnels de Cox (PHM), les réseaux de neurones artificiels (ANNs) et les machines à vecteurs de support (SVMs). Deux ensembles de données ont été utilisés pour illustrer la performance des trois méthodologies: l'ensemble de données du turboréacteur à double flux (turbofan) qui est disponible au sein de la base de données pour le développement d'algorithmes de pronostic de la NASA, et un autre ensemble de données obtenu d’une véritable application dans l'industrie. Les résultats de ces comparaisons indiquent que chacune des méthodologies proposées permet de prédire avec précision la RUL du système considéré.

Cette recherche doctorale conclut que l’approche utilisant la LAD possède d’importants mérites et avantages qui pourraient être bénéfiques au domaine du pronostic en CBM. Elle est capable de gérer les données en CBM qui sont corrélées et variantes dans le temps. Son autre avantage et qu’elle génère un savoir interprétable qui est bénéfique au personnel de maintenance.

# ABSTRACT

Estimating the remaining useful life (RUL) for a system working under different operating conditions represents a big challenge to the researchers in the condition-based maintenance (CBM) domain. The reason is that the relationship between the covariates that represent those operating conditions and the RUL is not fully understood in many practical cases, due to the high degree of correlation between such covariates, and their dependence on time. It is also difficult or even impossible for the experts to acquire and accumulate the knowledge from a complex system, where the failure of the system is regarded as the result of interaction and competition between several failure modes.

This thesis presents systematic CBM prognostic methodologies based on a pattern-based machine learning and knowledge discovery approach called Logical Analysis of Data (LAD). The proposed methodologies comprise different implementations of the LAD approach combined with non-parametric reliability estimation methods. The objective of these methodologies is to predict the RUL of the monitored system while considering the analysis of single or multiple failure modes. Three different methodologies are presented; two deal with single failure mode and one deals with multiple failure modes. The two methodologies for single mode prognostics differ in the way of representing the data.

The prognostic methodologies in this doctoral research have been tested and validated based on a set of widely known tests. In these tests, the methodologies were compared to well-known prognostic techniques; the proportional hazards model (PHM), artificial neural networks (ANNs) and support vector machines (SVMs). Two datasets were used to illustrate the performance of the three methodologies: the turbofan engine dataset that is available at NASA prognostic data repository, and another dataset collected from a real application in the industry. The results of these comparisons indicate that each of the proposed methodologies provides an accurate prediction for the RUL of the monitored system.

This doctoral research concludes that the LAD approach has attractive merits and advantages that add benefits to the field of prognostics. It is capable of dealing with the CBM data that are correlated and time-varying. Another advantage is its generation of an interpretable knowledge that is beneficial to the maintenance personnel.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF ABBREVIATIONS AND NOTATIONS

**Abbreviations**

| | |
|---|---|
| AHM | Asset Health Management |
| RCM | Reliability Centred Maintenance |
| CBM | Condition-Based Maintenance |
| IVHM | Integrated Vehicle Health Management |
| IAEA | International Atomic Energy Agency |
| RUL | Remaining Useful Life |
| LAD | Logical Analysis of Data |
| NASA | National Aeronautics and Space Administration |
| OSA-CBM | Open System Architecture for CBM |
| SNR | Signal-to-Noise Ratio |
| PCA | Principal Component Analysis |
| FDA | Fisher Discriminant Analysis |
| ICA | Independent Component Analysis |
| RMS | Root Mean Square |
| TSA | Time Synchronous Averaging |
| ARMA | Autoregressive Moving Average |
| FFT | Fast Fourier Transform |
| STFT | Short Time Fourier Transform |
| WT | Wavelet Transform |
| DWT | Discrete Wavelet Transform |
| WPT | Wavelet Packet Transform |
| HWPT | Harmonic Wavelet Packet Transform |

| | |
|---|---|
| GA | Genetic Algorithm |
| SBS | sequential backward selection |
| DET | distance evaluation technique |
| KM | Kaplan-Meier |
| MRUL | Mean Remaining Useful Life |
| PHM | Proportional Hazards Model |
| CIF | Cumulative Incidence Function |
| PNs | Petri Nets |
| ANNs | Artificial Neural Networks |
| WMRs | Wheeled Mobile Robots |
| FEA | Finite Element Analysis |
| PF | Particle Filtering |
| EPGS | Electrical Power Generation and Storage |
| EKF | Extended Kalman Filtering |
| HMM | Hidden Markov Model |
| SVR | Support Vector Regression |
| DWNN | Dynamic Wavelet Neural Network |
| PSD | Power Spectral Density |
| GRNN | Generalized Regression Neural Network |
| RNN | Recurrent Neural Networks |
| FCRNN | Fully Connected Recurrent Neural Network |
| EHSVs | Electrohydraulic Servo Valves |
| FFNN | Feed-Forward Neural Network |
| DGDS | Hazardous Gas Detection System |

| | |
|---|---|
| MCS | Monte Carlo Simulation |
| LR | Logistic Regression |
| SVMs | Support Vector Machines |
| SLT | Statistical Learning Theory |
| OAO | One-Against-One |
| OAA | One-Against-All |
| DAG | Direct Acyclic Graph |
| HVAC | High Voltage AC Machines |
| MILP | Mixed Integer and Linear Programming |
| SCP | Set Covering Problem |
| LASD | Logical Analysis of Survival Data |

**Notations**

| | |
|---|---|
| $S(t)$ | The survival function at time $t$. |
| $d_j$ | The number of systems that failed at time $t_j$. |
| $n_j$ | The number of systems which are at risk just before the time $t_j$. |
| $MRUL(t_k)$ | The mean remaining useful life calculated at time $t_k$. |
| $E[T - t_k \| T > t_k]$ | The expected value of the conditional random variable $T - t_k$, given the time $t_k$, where $T$ is the time to failure. |
| $Cov(p)$ | The set of observations covered by the pattern $p$. |
| $\Omega^{SL}$ | The set of observations collected from the short life systems that fail before a specified time $t_S$. The time $t_S$ is specified and decided by the maintenance personnel. In this thesis we consider $t_S$, the mean time to failure. |

| | |
|---|---|
| $\Omega^{LL}$ | The set of observations collected from the long life systems that fail after the time $t_S$. |
| $\Omega$ | The training data set. |
| $O(i, t_k, Z_{i,t_k})$ | The observation collected from the $i^{th}$ system $(i = 1,2, \dots M)$ at time $t_k$ $(t_k = 1,2, \dots t_{Fi})$, where $t_{Fi}$ is the failure time. The vector of covariates $Z_{i,t_k}$ represents the operating conditions and condition indicators at time $t_k$. |
| $O(i, t_{Fi}, Z_{i,t_{F_i}})$ | The failure observation of the $i^{th}$ system at time $t_{Fi}$, and $Z_{i,t_{Fi}}$ is the corresponding vector of the covariates. |
| $O\left(i, t_{Fi} \leq t_S, Z_{i,t_{Fi} \leq t_S}\right)$ | A positive (short life) observation in the training dataset. |
| $O\left(i, t_{Fi} > t_S, Z_{i,t_{Fi} > t_S}\right)$ | A negative (long life) observation in the training dataset. |
| $O(u, t_k, Z_{u,t_k})$ | The updating observation collected from the $u^{th}$ system $(u = 1,2, \dots U)$ at time $t_k$ $(t_k = 1,2, \dots t_{Fu})$, and $Z_{u,t_k}$is the vector of covariates at time $t_k$. |
| $S_b(t)$ | The baseline survival function (also called survival probability function) estimated by KM. |
| $P$ | The set of generated patterns, in the first article of the thesis. |
| $\lvert P \rvert$ | The cardinality of the set of generated patterns $P$ (i.e. the number of generated patterns). |
| $p_j$ | A pattern belongs to the set of generated patterns $P$, where $j = 1,2, \dots \lvert P \rvert$. |
| $S_{p_j}(t)$ | The survival curve of the generated pattern $p_j$. |
| $S_{O(u,t_k,Z_{u,t_k})}(t)$ | The updated survival curve of the updating observation $O(u, t_k, Z_{u,t_k})$, at time $t_k$ for the $u^{th}$ system $(u = 1,2, \dots U)$, and $U$ is the number of systems in the updating dataset. |

| | |
|---|---|
| $S_f(t)$ | The former updated survival curve obtained from the previous updating observation. |
| $MRUL_u(t_k)$ | The mean remaining useful life of the $u^{th}$ system , calculated at time $t_k$. |
| $\Delta t_r$ | The monitoring interval which is the difference between two consecutive inspection times i.e. $\Delta t_r = t_{r+1} - t_r$. |
| $RMSE(u)$ | The calculated root mean squared error (*RMSE*) for the *MRUL* estimation of the $u^{th}$ system in the updating dataset, where $(u = 1.2, \dots U)$. |
| $N_{Fu}$ | The actual number of operational cycles until the failure of the $u^{th}$ system. |
| $F_r$ | Friedman test statistic. |
| $P(\chi^2_{k-1} \geq F_r)$ | The calculated significance level, where $\chi^2_{k-1}$ is the Chi-Square value with $k$ degrees of freedom, and $k$ is the number of prognostic formulas. |
| $\chi^2_{k-1,\alpha}$ | The Chi-Square value with $k$ degrees of freedom and a declared significance level $\alpha$. |
| $d_{\alpha_F}\sqrt{Uk(k+1)/6}$ | A post-hoc value used for the second phase of Friedman test, where $d_{\alpha_F}$ is the $100(1-\alpha_F)^{th}$ of the standard normal distribution, and $\alpha_F$ is the family wise significance level. |
| $p_g \in P_{Gen}$ | A pattern belong to the set of generated patterns $P_{Gen}$, in the second article of the thesis. |
| $\lvert P_{Gen} \rvert$ | The cardinality of the set of generated patterns $P_{Gen}$ (i.e. the number of generated patterns). |
| $p_s \in P_{Sel}$ | A selected pattern belonging to the set of selected patterns $P_{Sel}$. |
| $\lvert P_{Sel} \rvert$ | The cardinality of the set $P_{Sel}$ (the number of selected patterns). |
| $W_{p_s}$ | The normalized weight of the selected pattern $p_s$. |
| $cov(p_s)$ | The set of observations that are covered by the selected pattern $p_s$. |
| $S_{p_s}(t)$ | The survival curve of the selected pattern $p_s$. |
| $S(\tau, Z_\tau)$ | The survival function at time $\tau$ for a vector of covariates $Z_\tau$. |

| | |
|---|---|
| $MRUL_u(t_k, Z_{t_k})$ | The mean remaining useful life (*MRUL*) calculated for the $u^{th}$ system at time $t_k$ for a vector of covariates $Z_{t_k}$. |
| $E[T - t_k \mid T > t_k, Z_{t_k}]$ | The expected value of the conditional random variable $T - t_k$, given the time $t_k$ and the vector of covariates $Z_{t_k}$. |
| $P(T \geq t_k)$ | The probability that the random variable $T$ is greater than the value $t_k$. |
| $S_i(t)$ | The cause-specific survival (sub-survival) for the failure mode $i$, where $(i = 1, \ldots, C)$. |
| $q_i$ | The proportion of systems that failed due to the failure mode $i$. |
| $h_i(t_j)$ | The sub-hazard function for the failure mode $i$ at $t_j$. |
| $f_i(t_j)$ | The sub-density function for the failure mode $i$ at time $t_j$. |
| $d_{ij}$ | The number of systems that have experienced the failure mode $i$ at $t_j$. |
| $\hat{F}_i(t)$ | The estimated CIF for the system that is experiencing the failure mode $i$ at the time $t$. |
| $S_{O(t_k)}(t)$ | The updated survival curve for the system based on the collected observation $O$, at time $t_k$. |
| $MRUL_i(t_k)$ | The *MRUL* for the cause-specific failure mode $i$, calculated at time $t_k$. |
| $\hat{S}_i(t)$ | The estimated sub-survival function derived from the estimated CIF for the failure mode $i$. |
| $\widehat{MRUL}_{O(t_k)}$ | The $MRUL$ of the monitored system estimated based on the updated survival curve $S_{O(t_k)}(t)$. |
| $\bar{E}$ | The average absolute prediction error. |
| $L_{Acc}(t_k)$ | The prediction accuracy function which measures the difference between the actual $RUL$ and the estimated $RUL$ at the inspection time $t_k$. |
| $L_{AP}(\mathrm{bearing}_g)$ | The average penalty incurred for the bearing $g$. |
| $L(\mathrm{model})$ | The total penalty incurred for each prediction model. |

# LIST OF APPENDICES

# CHAPTER 1 INTRODUCTION

Asset health management (AHM) system implies the continuous monitoring of an asset, in order to detect and predict as early as possible any changes in its condition that may have drastic consequences [1]. These consequences may concern safety issues, e.g. the failure of nuclear reactors or aircraft gas turbine engines. These consequences may also be economical when dealing with the occurrence of unscheduled downtime that may prevent accomplishing the committed mission schedules. As an example, unscheduled downtime has serious economic consequences in critical applications such as aircrafts or missiles assigned to space exploration missions [2]. These consequences provide the motivation to perform maintenance and repair tasks before such drastic situations arise.

The task of maintenance is to maintain reliable and cost-effective operations of physical assets. Different maintenance strategies have been introduced as an efficient way to assure a satisfactory level of availability of the assets [3]. Moreover, the maintenance of an engineering asset is a prerequisite which has significant impacts on its safety and reliability [4].

Maintenance costs have increased rapidly during the past years since they represent a major part (from 15% to 40%) of the total operating costs in manufacturing and production plants, according to the study reported in [4, 5]. It was also reported that as much as 30% of the total maintenance expenditures were spent on some unnecessary actions like, the use of unnecessary preventive maintenance, and wrong scheduling [6]. Consequently, the experts in maintenance engineering have paid a deep attention to various types of advanced maintenance strategies [7].

## 1.1 Taxonomy of maintenance strategies

Various definitions of maintenance strategies have been suggested during the past years. The maintenance terminologies used in this thesis have been used in [4, 8, 9]. In a broad sense, the maintenance strategies can be classified as reactive (also called unplanned or break-down or run-to-failure maintenance) and proactive (also called planned maintenance). Figure 1-1 gives a brief taxonomy for different maintenance strategies, as presented in [2].

In reactive maintenance, the actions are taken after breakdowns occur (fix or replace the unit after it fails). In proactive maintenance, the objective is to resolve the issues in the operating asset prior to the onset of failure [10]. This can be done by performing inspection and/or servicing tasks that have been pre-planned to retain the functional capabilities of the asset [11]. Proactive maintenance strategies can be further classified into preventive and predictive maintenance. In preventive maintenance, the actions are taken in predetermined or periodic intervals in order to prevent breakdowns or failures, regardless of the health state of the physical asset [9].

As technology developed rapidly, engineering systems have become more and more complex while higher quality and reliability are required [12]. As a result, the costs of preventive maintenance become higher. One way to minimize unnecessary maintenance and repair costs and probability of failure is to perform an assessment and prediction of the asset's health state, based on its current and historical operating conditions [2]. This can be carried out by applying a predictive maintenance strategy where the action taken is dependent on the nature of the physical asset and the availability of input data.

In predictive maintenance, the goal is to make economically justifiable decisions by blending together the economic aspects and the risk prediction, in order to identify the optimal maintenance decision [13]. The authors in [2] present a good review for the maintenance strategies that focus on improving reliability and reducing unscheduled downtime by monitoring and predicting the asset's health conditions. As depicted in Figure 1-1, predictive maintenance can be further classified into reliability centered maintenance (RCM) and condition-based maintenance (CBM).

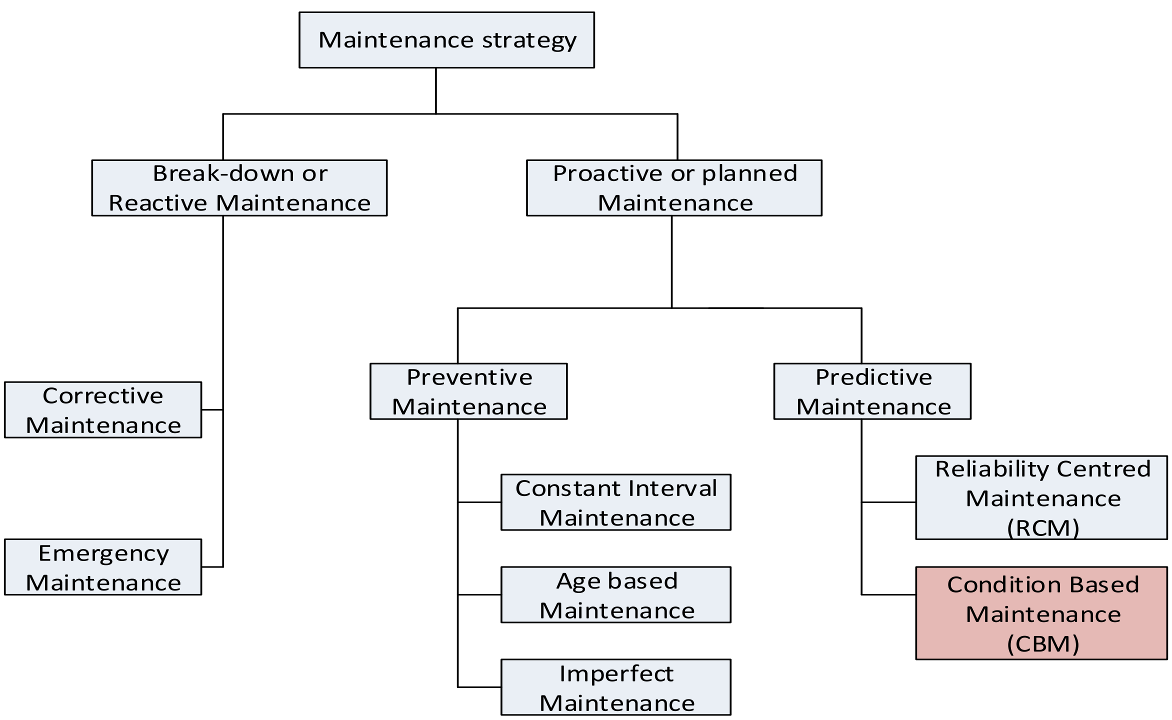


Figure 1-1: Taxonomy of maintenance strategies as presented in [2]

Recently, CBM has received an increasing attention among the practitioners and researchers, as a predictive maintenance strategy [5]. It consists of activities and tasks for detecting and correcting failure causes, by monitoring the symptomatic conditions of the failure process in the targeted asset. CBM utilizes condition monitoring technologies to detect and predict the current and future health states of the engineering assets, based on non-intrusive measurements of their operating conditions [14].

CBM strategy attempts to avoid unnecessary preventive maintenance tasks by taking maintenance actions only when there are abnormal signs coming from the monitored asset [15]. In other words, it makes a maintenance decision based on the up-to-date degradation data of the monitored asset. Many cases from the industry that are applying CBM show significant reductions in preventive maintenance costs, while maintaining or even improving the availability and reliability of the monitored systems [13].

AHM system uses the CBM strategy to provide an intuitive and integrated solution to detect the defects of an asset through the continuous monitoring during multiple degraded states before it fails [1]. In the AHM/CBM system, the degraded states of the asset are monitored and predicted, and the optimal maintenance actions can be taken to protect it from breakdown or catastrophic failures [3].

Researches supported by international agencies, industry and academia are focusing on designing more effective and intelligent AHM/CBM systems [16]. The National Aeronautics and Space Administration (NASA) uses the Integrated Vehicle Health Management (IVHM) program for its fleet [17]. The International Atomic Energy Agency (IAEA) is continuously paying a great attention to AHM/CBM programs, in order to improve both reliability and safety of nuclear power plants [18].

As decision making procedures in the CBM system, diagnostics and prognostics are two important aspects. Diagnostics deals with fault detection, isolation and identification when it occurs [3, 19, 20]. Fault detection is the task of indicating whether something is going wrong in the monitored asset or not; fault isolation is the task of locating the faulty component; and fault identification is the task of determining the nature of the fault in the located faulty component [3, 20].

Prognostics on the other hand deal with the prediction of failure before it occurs. The task of such prediction is to determine whether a failure is impending and estimate how long and how likely this failure will occur [20]. A particular prognostic issue is to estimate the remaining useful life (RUL) (also called residual life) of the monitored asset. The RUL is defined as the time left before the occurrence of catastrophic failure [12, 20].

Developing an accurate prognostic method is motivated by the need to prolong the utilization of the system. The accuracy of diagnosis procedure affects the life of the monitored system since the prognosis procedure is based on these diagnostic results [3]. Therefore, the prognosis procedure requires precise diagnostic models, in order to estimate the future health states and the RUL of the monitored system, accurately [21].

Fortunately, the invention of advanced sensors and powerful signal processing techniques enable the maintenance practitioners to extract multiple features for the purpose of degradation detection and quantification [22]. Some of the extracted features can be selected and can serve as the basis for accurate diagnosis and hence RUL prediction. The prediction of RUL in CBM is not completely understood in many practical situations [20]. As an example in bearing prognostics, it is known that the vibration in a pump is not the only indication for its future failure.

The aim of this thesis is to present the subject of this doctoral research which focuses on the implementation of an interpretable knowledge discovery approach called Logical Analysis of Data (LAD) in the field of CBM prognostics. LAD was applied successfully in the field of medical diagnosis and prognosis to diagnose patient’s condition and to predict the propagation of some diseases [23-25]. As a machine learning classification technique, LAD was used on a great variety of problems and reacts well to noisy data and measurement errors [26]. It was applied in the field of CBM diagnostics for the first time by the maintenance and reliability research team at École Polytechnique de Montréal, Canada [27].

The material discussed and the terminologies used in this thesis are often interdisciplinary, originating in quite disparate fields such as machine learning, pattern recognition, statistical models, mechanical engineering, computer science, and maintenance and reliability engineering. The terms used in this thesis are largely those pertaining to the CBM domain and are taken from some of the most cited papers and books such as [3, 12, 20]. Other terminologies from different disciplines may be brought to the CBM domain. The vocabulary used in Chapter 2 explaining the CBM and LAD are adopted from [3, 26], as they are the most cited articles concerning the two domains.

## 1.2 Problem statement

A significant amount of researches and reasonable progresses have been achieved in system diagnostics and it has been a subject of considerable attention in CBM community, whereas the prognostic methods have not enjoyed the same attention [3]. In general, the design procedures for the prognostic methods are more complex than that of the diagnostic ones, due to some challenges and complications. As the systems become complex and critical, the need for accurate prognostic methods has been urgent in order to maximize their reliability. Maximizing the reliability of a system means increasing its ability to survive more and more in the future [28]. An example of such systems is the aircraft turbine engine [29].

In order to compare different prognostic methods, some performance indices should be followed [20]. A justification of a designed prognostic method should be provided based on such performance indices [12]. From data analysis point of view, as more historical data about the system's health conditions become available, the devised prognostic method will be able to predict the RUL more accurately [20].

We identified two main research problems in the field of prognostics, they are stated as follows:

1- A significant challenge is to estimate the RUL for a population of systems working under varying operating conditions. This is because the relationship between the covariates that are representing those operating conditions and the prediction of RUL in many cases is not fully understood, due to the high degree of correlation between such covariates and their dependence on time. Generally, long term prediction of RUL in CBM entails large amount of uncertainty because the system's degradation undergoes dynamic stochastic process and usually consists of a sequence of degraded states [22].

2- Another challenge is to estimate the RUL in case of complex systems that fail due to one of multiple failure modes. In this case, the failure of the system is regarded as the result of interaction and competition between several failure modes. In rotating machinery as an example, at the system level, different components can have different failures such as: bearing defect, cracked or broken rotor bars, mechanical seal wear, and others [30]. At the component level, there may be different types of failure modes such as: an inner race defect, an outer race defect, a crack in the cage of the rolling element. In multiple failure modes prognostics, the challenge is to consider the effect of those interacting and competing failure modes when analyzing the RUL or the failure time of a monitored system.

## 1.3 Limitations of current prognostic methods

Most of the current prognostic methods in the CBM literature do not address the above two challanges more accurately. Accordingly, they pose some limitations summarized in the following two papagraphs.

The majority of those prognostic methods either assume many impractical assumptions about the disribution of the data, or they are based on tedious parameter tuning and experimental settings. A number of these prognostic methods still rely on human experience. Although the experts have significant knowledge about system failure and its degraded states, they do not have systematic methodologies that can predict the RUL in the presence of highly correlated or time-varying covariates [2]. Therefore, there is a need to develop and to improve systematic prognostic methodologies that can be implemented in the AHM/CBM system without the need to hire an expert.

In case of multiple failure modes prognostics, most of the prognostic methods assume that the lifetime data for each failure mode are drawn from a certain parametric failure distribution function. There are two major limitations when applying such parametric prognostic models in CBM. The first limitation of such models is that they are application-dependent. This means that the assumption that allows the lifetime data for each failure mode to follow a certain distribution, needs a lot of experience and knowledge about the application at hand. It is also hard or even impossible to accumulate the knowledge in the case of complex systems that have a multitude of failure modes. The second limitation of those prognostic models is that the failure distribution function of each failure mode is estimated by considering the other competing failure modes as censored categories. According to those two limitations, it is necessary to develop a non-parametric failure distribution function for each failure mode, while considering the other failure modes as competing ones.

The above problems and limitations in CBM prognostics motivate us to design and develop prognostic methodologies that are not based on any assumptions for the distributions of CBM data, and are not based on any experts as well. The methodologies are based on certain theoretical and practical settings, targeting some issues of implementing CBM prognostics as a holistic phenomenon. They are based on LAD as a proven approach which possesses two key advantages over the other knowledge discovery techniques.

## 1.4 The advantages of using LAD as knowledge discovery approach

The results reported in the CBM diagnostic applications presented in [10, 27, 31, 32], show that LAD as a classification technique demonstrated a good performance in detecting the faulty systems. It is also concluded from the results presented in those works that LAD is a promising tool in CBM decision making. Moreover, it is shown that the LAD approach has some important advantages over the other comparable techniques. These advantages are listed as follows:

1- Unlike many CBM decision making techniques which assume that the input data belong to a certain probability distribution, LAD is not based on any statistical analysis. This makes it capable of dealing with the covariates that are highly correlated and time-varying, without the need to satisfy any statistical assumptions. LAD detects and evaluates the correlation between any number of covariates in the CBM data, without the need to select the most significant ones.

2- The transparency (interpretability) and knowledge preservation. This is one of the important properties of the LAD approach when it is compared to the other machine learning techniques in the field of CBM [32]. The transparency means that LAD as a decision model is meaningful and gives a physical interpretation about the behavior of the monitored system. Once the training phase is completed, the knowledge is discovered in the form of interpretable patterns and preserved for the future use by the maintenance personnel.

## 1.5 Research objectives

### 1.5.1 General objective

The ultimate objective of this doctoral research is to estimate the future health condition and to predict the failure time of the monitored systems using automated CBM prognostic methodologies based on LAD.

The task is to predict the advent of failure in terms of the probability of mission survival or remaining useful life. The historical CBM data collected from various systems are analyzed for the purpose of knowledge discovery. The discovered knowledge is then used in different ways by using diagnostic and prognostic algorithms.

In this doctoral reseach, different implementation methodologies are adopted by combining LAD to non-parametric reliability estimation methods. These methodologies will be discussed in the subsequent chapters of this thesis. The developed prognostic methodologies can be employed to reduce industry's dependence on the experts.

The performance of each methodology in this doctoral research is measured in terms of the difference between the predicted and actual RUL of the monitored system.

### 1.5.2 Specific research objectives

The above general objective is achieved by fulfilling the following specific objectives:

1- Implementation of two different prognostic methodologies based on the two-class LAD approach. The two methodologies are developed in order to predict the RUL of a system works under varying operating conditions. LAD is exploited to extract the hidden knowledge from the CBM data. The extracted knowledge is then used to update the reliability curve estimated by using a common non-parametric estimation method. The updated reliability curve is then exploited to estimate the system's RUL.

2- Implementation of multi-class LAD as a prognostic methodology for the systems that fail due to multiple failure modes. The multi-class LAD is merged with a set of non-parametric functions that are characterizing the different failure modes (one function for each failure mode). The reliability curve of a monitored system is updated based on the knowledge extracted from the CBM data. The RUL is then estimated based on the updated curve.

3- Software development. The prognostic algorithms in the three methodologies are developed using designed computer codes. The codes are integrated and linked to the existing software *cbmLAD*, to deal with the CBM prognostic applications. The intellectual property of the developed software belongs to École Polytechnique de Montréal, and those members of our maintenance and relibility research team who contributed in the development of this software.

## 1.6 Originality and success

The maintenance and reliability research team at École Polytechnique de Montréal, Canada was able to reproduce human expertise in detecting and analyzing different faults in the field of CBM diagnostics, by using LAD as a classification method. LAD was used successfully to diagnose machine conditions and to detect faults of some important applications such as rotor bearings and power transformers. It also demonstrated good performance in detecting and analysing the faults and phenomenon of rogue components in airplanes. It is clear that the accuracy of a prognostic algorithm is mainly dependent on the accuracy of diagnostic information used. Our hypothesis states that LAD diagnostic model can be developed and used as the cornerstone in different prognostic methodologies to predict the RUL. According to the previously mentioned advantages of LAD, this thesis presents novel CBM prognostic methodologies based on LAD. The novelty of this doctoral research is stated in two points (1) the application of two-class LAD as prognostic methodology, and enhancing this methodology in order to improve its performance (2) A novel and innovative application of multi-class LAD within the context of multiple failure modes prognostics in CBM. To the best of our knowledge, the material presented in this thesis is not published or written elsewhere except where due references are cited.

The methodologies involved in this thesis were compared to other common prognostic methods in the CBM domain. The obtained results are promising and indicate that these methodologies can be applied successfully in the industrial applications.

## 1.7 Deliverables

The research conducted in this thesis was followed in an evolutionary manner in order to achieve the objectives stated previously. The outcomes of this doctoral research are:

1- Three conference papers.

2- Three articles, two focus on the implementation of two-class LAD in prognostics and the third one focuses on the implementation of multi-class LAD in multiple failure modes prognostics.

3- A thesis compiling the findings, including the three articles.

4- In addition to the articles, a related software was developed.

This doctoral research is documented in the three articles incorporated into this thesis. The first two articles address the first problem stated previously. The first article in this thesis has been approved and published in the “Journal of Intelligent Manufacturing”. The second article addresses three modifications to the methodology proposed in the first article. It has been submitted to the journal of “IEEE Transactions on Reliability”.

The third article addresses the second problem stated above. It involves the implementation of multi-class LAD as a multiple failure modes prognostic methodology in a common CBM appliaction; rotating machniery prognostics. It has been submitted to the journal of “Mechanical Systems and Signal Processing”. Figure 1-2 illustrates the evolutionary approach followed in this doctoral research along with the three articles involved.

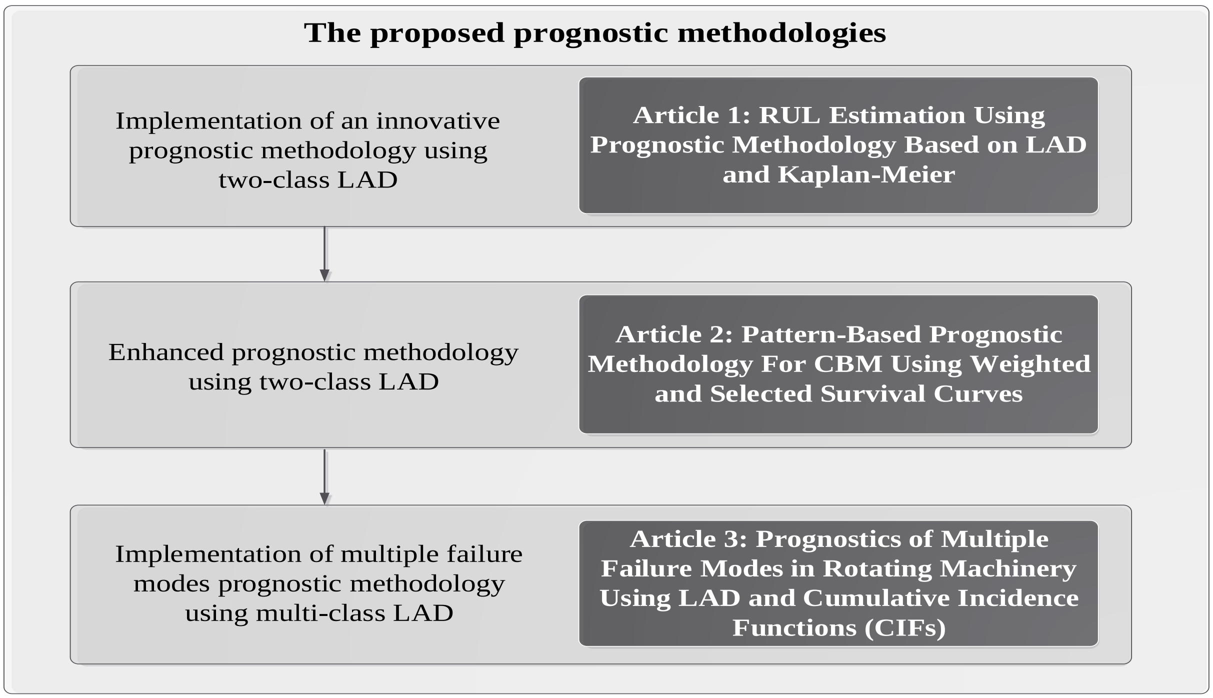


Figure 1-2: Illustration of published and submitted articles incorporated in this thesis

## 1.8 Social impacts and economic benefits

The results obtained during this doctoral research were published and submitted in international journals and specialized conferences, targeting to enforce the link between both academic and industrial communities. The results obtained were validated in collaboration with selected practitioners in North America such as NASA and Prüftechnik Canada. This research on the other hand is expected to reduce the maintenance cost significantly since it is able to accurately predict the system's health conditions. Hence, it enables the maintenance personnel to act only when maintenance is actually necessary. We are planning to extend this research in the future to the optimization of maintenance resources which is one of the interests of our research team at École Polytechnique de Montréal.

## 1.9 Thesis organization

This thesis is divided into six chapters. Chapter 1 explains the problem being studied and its scientific, practical, and economical pertinence. It also states the objective of this research and explains its originality, the targeted deliverables, and delimitations of the thesis.

Chapter 2 explains the different procedures involved in CBM and the techniques used in each procedure. It also presents a literature review on the current CBM prognostic methods. The purpose is to give the reader a general overview and a preliminary number of references for the common prognostic methods. Meanwhile, the chapter presents the LAD decision model and describes the state of the art in this approach. The historical perspectives and knowledge discovery process are overviewd. The common applications of LAD are presented at the end of the chapter.

Chapters 3 through 5 present the three articles incorporated into this thesis. They include the developed prognostic methodologies used throughout this doctoral research.

Chapter 3 proposes a novel and innovative prognostic methodology to predict the RUL of a system working under different operating conditions, based on two-class LAD. The performance of the proposed methodology is illustrated on a benchmark dataset in the field of prognostics, by comparing it to the most common statistical-based CBM prognostic method; the proportional hazards model (PHM).

Chapter 4 presents an enhanced prognostic methodology addressing three modifications to the proposed methodology in Chapter 3. The utility of the proposed methodology is demonstrated through a number of experiments on the same benchmark dataset, by performing comparisons with two common machine learning techniques; support vector machines and artificial neural networks.

Chapter 5 presents a novel prognostic methodology based on multi-class LAD to predict the RUL of the systems that are working under difffernt operating conditions and subjected to multiple failure modes. The methodology addresses the two main limitations of the current multiple failure mode prognostic methods that were stated previously. The proposed methodology is validated through experiments on a dataset collected from the industry. It is compared to support vector machines and artificial neural networks, for the purpose of validation.

Chapter 6 presents a general discussion on the three articles incorporated in this thesis, followed by conclusions and future work directions to extend this doctoral research.

# CHAPTER 2 BACKGROUND AND LITERATURE REVIEW

## 2.1 CBM definition and standards

The definition of CBM in this thesis is adopted from [3] as one of the most cited papers in this domain. It is defined as the predictive maintenance initiated as a result of continuous or periodic measurement and interpretation of data to indicate the condition of the monitored asset [3, 33, 34]. The standard definition found in the British Standard (BS 3811:1984) defines the CBM as "*the preventive maintenance initiated as a result of knowledge of the condition of an item from routine or continuous monitoring* ", as stated in [35].

CBM is needed to guarantee the survival of a system so that incipient faults can be detected and diagnosed as early as possible. Although the possibility of failure occurrance cannot be avoided, the earlier diagnosis of an incipient fault is useful to avoid the occurrance of catestrophic failure in the system. When such fault exists, it will give some indicators (symptoms) such as excessive vibration and noise, increased temperature, oil debris, and others. These indicators carry valuable information about the state of the system, and they form the basis for CBM policy planning [36].

The technical constituents and organizational aspects in the CBM system are investigated in [37, 38]. The objective is to make the CBM strategy more accessible within the industry, and to increase technology modularization and system flexibility as well. Several standards and standardization reports were already published and have been available for the developers and customers of CBM system technology [38]. An Open System Architecture for CBM (OSA-CBM) for the organizations was presented in [39]. The architecture comprises seven layers as depicted Figure 2-1 [37]. This doctoral research focuses mainly on the fifth layer (prognostics) in this architecture.

Another version of an OSA-CBM has been presented recently in [40]. In that paper, the architecture was developed to address the need for a certain standard that handles the flow of information between the different components of the CBM system. Generally, the OSA-CBM exploits the advancements in the information and database technologies. As a consequence of utilizing such technologies, a huge volume of data are available and stored in the CBM databases. The analysis of such massive data creates new challenges to researchers and practitioners in the CBM domain.

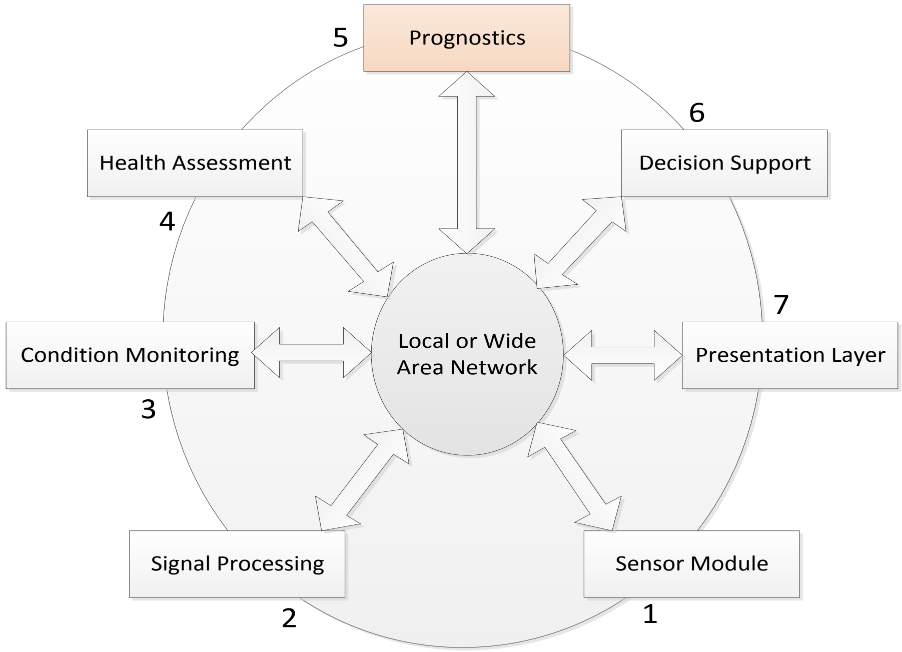


Figure 2-1: The seven functional layers of the OSA-CBM presented in [39]

## 2.2 Types of CBM data

The acquired data in the CBM system can be categorized into two main categories [3]: condition monitoring data and event data. Condition monitoring data are collected and processed to determine the system's health condition, where the measured observations are related to the health states of the monitored system [3]. If the condition monitoring data have trends that are reflecting the degradation conditions of the system, they are called degradation data [41]. They can represent the vibration, temperature, pressure, moisture, humidity, oil debris, and others.

Event data provide the information on what happened (failure, installation, overhaul, etc.) and/or what was done (repair, preventive maintenance, oil change, etc.) to the targeted system [3]. If the event data represent the failure of the system, they are called lifetime data.

**Types of condition monitoring data**

Condition monitoring data are further categorized into three subcategories: waveform data, value data, and multidimensional data [3]. Waveform type data are time series measurements which are collected over a certain time interval (time record), hence they are called time-domain data. Vibration and acoustic signals are common examples of waveform type data [42]. Value type data can be measured directly from the data acquisition module at a specific time instant or through processing the waveform data. Examples of this type of data are measures of temperature, pressure, humidity, etc. Multidimensional type data are measured at specific time instants over three or more features (dimensions). An example of such data is the infrared thermographic image [3].

In this doctoral research, the waveform data and value type data are the most relevant.

## 2.3 Data Mining and Knowledge Discovery in CBM

### 2.3.1 Big Data in CBM

There are two main problems associated with the existence of huge volumes of CBM data. First, it is reported in [43] that the amount of stored databases doubles every 20 months. Accordingly, it is difficult and sometimes impossible to deal with such amount of data in any quantitative sense (this is referred to as *Big Data*) [43]. Instead, the important information should be extracted from the data. Second, the CBM databases have the potential to predict the evolution of interesting events in terms of variables or trends which have not been fully exploited yet, because these huge data are not well represented.

The advancement of the software, hardware and methodological researchs have recently led to the development of flexible procedures that can be used to analyze the CBM data [9, 32, 44-46]. However, there is a need to develop a knowledge discovery procedure to deal with the CBM data in a systematic and clear way, in order to support the CBM maintenance personnel with an appropriate assessment for the health states of the monitored system.

### 2.3.2 Data Mining and machine learning in CBM

*Data mining* is defined as the process of automatic exploration and extraction of *knowledge* from the data [47]. *Machine learning* provides the technical basis of data mining. In other words, it is the technology for mining knowledge from data [43]. It relies on the availability of data and draw on learning strategies from the area of computational intelligence, pattern classification, and others [48]. The machine learning techniques are based on the concepts of training and testing.

In the context of CBM, the main objective of machine learning is to build computer programs that refine the CBM databases automatically in order to extract useful information. Then, the extracted information are then exploited to discover a set of patterns. Those patterns are representing a valuable knowledge about the history of the system's health states. Among the extracted patterns, some are trivial and nonsignificant, others are general and can contribute to an accurate prediction for the targeted events [43]. The patterns discovered should be meaningful and have some advantage in an economic sense. From economical point of view, one of the most important requirements for the patterns is the comprehensiveness (interpretability) [32]. Some patterns are comprehensible (also called transparent or interpretable or structural), while some are comprehensive but not necessarily comprehensible (called black box patterns) [49]. From the performance point of view, both of them can have good prediction [48].

The advantage of using the interpretable patterns in CBM is their transparency, therefore they can help the analyst and the decision maker to explain the predicted events in an explicit way.

Based on the extracted patterns, the machine learning technique maps the CBM data into a decision model, in order to produce predicted outcomes for the new observations that are not found previously in the data. Those observations are collected recently from the monitored system, and they are carrying information about its health state. The decision model can be used as either a regression or classification technique, depending on the variable of interest to be predicted.

Machine learning classification techniques are further divided into supervised and unsupervised. In supervised learning, the purpose is to infer a decision model from labeled data. In unsupervised learning, the learning technique is fed with only unlabeled observations (there is no a priori label).

## 2.4 CBM Components

The seven layers of the OSA-CBM depicted in Figure 2-1 can be aggregated into three main procedures in the CBM [3, 20]. In this thesis, we discuss the three aggregated procedures of the CBM program presented in [3]; namely:

1. Data acquisition.
2. Data processing.
3. Maintenance decision-making (diagnosis and prognosis)

### 2.4.1 Data acquisition

Data acquisition is the first procedure in the CBM program. In this procedure, the data are collected by the means of transducing devices, in their raw forms. The data must be refined appropriately in order to obtain proper measurements. The refined data will serve as the main input to the data processing procedure in the CBM program. The type of data processing algorithms and tools are dependent on the type of acquired data.

**Sensing strategies**

Sensors and sensing strategies constitute a major technology and foundational basis in CBM. They are intended to monitor typical state variables as vibration, temperature, pressure, speed, etc. Such variables are commonly used in CBM fault diagnostic/prognostic of machinery [20]. In the case of rolling elements diagnosis/prognosis for example, vibration data are needed because they relate the trends of vibration features to the degradation and failure process of bearings [50, 51].

More specifically, some sensors are designed to measure quantities that are directly related to the failure modes which are candidates for diagnosis and prognosis [20]. Common examples of such sensors are: ultrasonic sensors, strain gauges, acoustic emission sensors, electrochemical fatigue sensors, and proximity devices [52-54]. Other sensors are designed as multipurpose sensors to monitor process variables in CBM [20]. Recently, wireless sensors have been introduced in CBM domain [55-57].

**Signal conditioning**

Signal conditioning is an important and required step in data acquisition procedure. The signals coming from the sensors can be very noisy, of low amplitude, and dependent on secondary variables such as temperature [20]. As a consequence, we may not be able to measure the quantity of interest but only the dependent quantity. The sensor output must be conditioned and processed to provide an appropriate measurement for the physical quantity [58]. Amplification, level translation, linearization, and filtering are fundamental signal conditioning tasks [59]. A common objective of signal conditioning is to improve the signal-to-noise ratio (SNR), in order to extract as useful as possible information from the raw data. Signal conditioning may be carried out with hardware or software. There is a variety of integrated circuits available for hardware signal conditioning. The signal conditioning in software is carried out much more efficiently and accurately, and this eliminates the need for difficult hardware calibrations [58].

### 2.4.2 Data processing

Data processing is the second procedure in the CBM [3]. In this procedure, the data acquisition module feeds the CBM system with the acquired signals, in order to process and extract the relevant features as a first step [60]. The second step in data processing procedure is the feature selection step. In feature selection step, the purpose is twofolds, firstly, to select the optimal number of features that reduces the dimensionality of feature space and secondly, to select the superior features that improve the performance of the CBM decision making procedure [48].

**a-Feature extraction**

Generally, it is difficult to assess the health states of complex systems directly from the measured time-domain data. Those data are collected by the data acquisition module and are not useful in their raw forms. The acquired data must be preprocessed appropriately, to extract useful information that represent a reduced version of the original data, but preserve as much as possible some characteristic features [48, 61]. In the practice, it is useful to extract some features from these time-domain data as good representatives of state changes in the system being monitored [31]. These features are the indicators for the faulty events we are seeking to detect, isolate, and predict [20].

For multidimensional type data, feature extraction using image processing is more complicated than that of waveform signal, due to the higher dimensionality involved. The resulting feature space after processing an image is a high dimensional space, since it represents all the constituent pixels in the image. Increasing the dimensionality of such space makes the training data sparse. This is a common problem in pattern recognition literature, it is known as *curse of dimensionality* [61]. The higher dimensionality leads to poor generalization (known as overfitting problem), and therefore decreases the performance of the decision model [48, 61].

One way to deal with such problem is to apply some processing techniques such as: principal component analysis (PCA) [62], Fisher discriminant analysis (FDA) [63], and independent component analysis (ICA) [64], to reduce the dimensionality of the feature space, while keeping the classification accuracy [48, 61]. In the context of CBM, other examples of processing techniques are presented in [65, 66].

In case of waveform type data, processing techniques such as time–domain analysis, frequency–domain analysis, and time–frequency analysis are the most applicable [3]. Since this type of data is involved in this doctoral research, descriptions of these analysis techniques will be discussed briefly in this thesis, in the following.

**Time-domain analysis**

Time-domain analysis techniques are used to reflect the statistical behavior of waveform signals. In these techniques, some commonly descriptive statistics or higher order statistics are calculated [67]. Examples of such descriptive statistics are: mean, peak, peak-to-peak, standard deviation, root mean square (RMS), and crest factor [68]. Examples of higher order statistics are: kurtosis and skewness [69]. The extracted features are called statistical features since they are based on the distribution of the waveform signal that is treated as a random variable. Vibration signals are the most common and popular examples of waveform type data in CBM [67]. The statistical features extracted from such signals are not effective measures when the acquired signal is a mixture of signals coming from more than one source of vibration, in the system being monitored [3].

Another popular time-domain analysis technique is the time synchronous averaging (TSA). The waveform signal is divided into segments of equal length based on a certain synchronous signal (e.g. a tachometer signal can be used as a synchronous signal in rotating machinery). The idea of TSA is to average the waveform signal over a number of segments, in order to remove or to reduce the noise, thus having an improved estimate for the desired signal [70]. One of the disadvantages of TSA when it is applied in rotating machinery prognostics, is its sensitivity to speed changes. If the speed of the rotating machine changes, the sample length and the number of points in each segment will vary accordingly [68]. Other approaches apply time series models to waveform data such as autoregressive (AR) and autoregressive moving average (ARMA) models [3]. The objective behind these models is to extract features by fitting the waveform data to a parametric model. The disadvantage of time-domain analysis techniques is their weak performance when applied to modulated non-stationary signals, which exist commonly in industrial machinery [31].

**Frequency-domain analysis**

Frequency-domain analysis techniques have been the subject of extensive research over many years [71]. In such techniques, the time-domain waveform is transformed to the frequency-domain using fast Fourier transform (FFT), for better identification of frequency components. It is also called spectrum analysis. The objective is to inspect the whole or certain frequency components, thus extract features from the spectrum. In the context of CBM, the frequency spectrum carries a great deal of useful information in the diagnosis and prognosis of rotating machinery [22]. The amplitudes of peaks found in the spectrum indicate the type of fault and its severity [20]. Other frequency-domain analysis techniques such as Cepstrum and Waterfall have been developed to detect the faults in CBM [67]. These methods have been shown to have their own advantages over spectrum analysis methods in some specific cases [3, 51].

The disadvantage of using the frequency-domain analysis techniques is their basis functions which are extending over an infinite period of time [71]. This limitation hinders these techniques to handle non-stationary transient signals with short durations [20]. Such transient signals come when the components of defect frequencies overlap each other in the machine [71].

**Time-frequency domain analysis**

As one of the most popular time-frequency techniques, short time fourier transform (STFT) was developed to address the limitation of the FFT [72]. The objective is to divide the whole waveform signal into a number of segments with a short time window and then apply the FFT to each segment. Some of the popular window functions are Hamming, Hann, and Gaussian functions [71]. The disadvantage of the STFT is that its time resolution and frequency resolution cannot be chosen to be simultaneously small, according to the Heisenberg's uncertainty principle [73]. A trade off must be made between the time resolution and frequency resolution.

Wavelet Transform (WT) was suggested as a multiscale time-frequency analysis for non-stationary signals through dilation and translation processes, aimed at extracting time-frequency features [72]. WT uses a series of oscillating functions with different frequencies as window functions (the mother wavelet) to scan and to translate the time waveform. The mother wavelet function is a zero average oscillatory function centered around the zero with a finite energy. It can take any possible form, provided that it satisfies the conditions of admissibility found in [74]. Some commonly used mother wavelets are the Morlet, Haar wavelet, Mexican Hat, Daubechies wavelets, etc [71]. At high frequencies, the wavelet reaches a higher time resolution but a lower frequency resolution, while at low frequencies; it reaches a lower time resolution but a higher frequency resolution [75]. This multiresolution characteristic is more advantageous for non-stationary signals with discontinuities and spikes.

The discrete wavelet transform (DWT) is presented as a discritized version of the WT to reduce redundancy that results from varying the scale and translation parameters continuously. It uses the discrete values of those parameters [75]. One of the advantages of the DWT is its ability to reduce noise in raw signals (denoising) [71].

As an extension for the DWT, wavelet packet transform (WPT) provides more flexible time-frequency decomposition. The WPT can decompose the waveform in high-frequency regions, thus allows feature extraction from sub-frequency bands of the decomposed signal, where the important features are hidden [71]. When the WPT is applied to vibration signals, not only all the major transient elements are identified, but also the corresponding frequency shifts [71].

Extensive researchs have been conducted through different settings of the WPT, in order to represent bearing vibration under different defect conditions [67, 71, 72, 76]. In [76], the discrete harmonic wavelet packet transform (DHWPT) was applied to vibration signals measured from rolling bearing, by decomposing each signal into a number of frequency sub-bands. Then, the key features associated with each sub-band were extracted.

There are several ways in which wavelet features can be extracted and applied in fault diagnosis and prognosis. Some of the most common extraction techniques for wavelet features are: coefficient-based, energy-based, singularity-based, and wavelet function-based methods [20].

**b-Feature selection**

Not all the extracted features are useful for the CBM decision-making procedure. The extracted feature set contains nonsignificant features and superior features as well. The usage of all the extracted features not only slows down the procedure of buliling a CBM decision-making model, but also tends to degrade its performance [20].

In CBM decision making, it is necessary to deal with the rich faulty information, thus superior features need to be selected from the original feature set, in order to obviously characterize the system’s conditions. The feature selection procedure further improves the decision-making procedure accuracy and reduces the computational burdens in the CBM program [20, 22, 77].

Many techniques have been proposed to perform feature selection. The distance evaluation technique (DET) [22, 78, 79] and genetic algorithm (GA) [80-82] were proposed for extracting the set of superior features. A feature selection technique employing sequential backward selection (SBS) is found in [83].

The DET is one of the most simple and efficient feature selection techniques in the field of CBM [22, 78, 79]. The idea of the DET algorithm is to select the features that make the distances within a set of classes shorter and the distances among classes longer. An evaluation ratio (weight) for each feature in the original feature set, is calculated. Based on these weights, all the features are ranked in an ascending order. According to a certain threshold value, the feature having a weight larger than the threshold is selected, otherwise it is discarded. The set of selected features are then applied to the decision-making procedure in CBM.

In [84], the features are extracted using PCA and ICA, then the superior features are selected using the DET. An improved versions of the DET are introduced in [77, 85, 86].

One of the important advantages of the DET over the GA and SBS methods is its tractability and excellent computational time. Another advantage of that technique over the PCA, FLD, and ICA is that it selects the set of features, without carriyng out any blind transformation for the original feature space. This makes the technique compatible with any decision-making model.

According to those two advantages, we use the DET algorithm in this doctoral research as a feature selection technique. The DET algorithm found in [77], is presented in Appendix A of this thesis.

### 2.4.3 Maintenance decision-making procedure (diagnosis and prognosis)

Decision-making is the third and final procedure in the CBM system presented in this thesis. The objective is to determine the optimal time to replacement or overhaul the components of the monitored system [2]. This can be performed based on a prediction or assessment of the system's health conditions and its residual life. The decision made based on accurate prediction is crucial to reduce down time, prolong the system's lifetime, improve productivity, and enhance its safety [87].

In CBM, diagnostics and prognostics are essential tools to identify incipient faults and to predict the RUL before catastrophic failure occurs [3]. They are the two main aspects to support maintenance engineers by the decision making capability [20].

Three main tasks can be performed by the diagnostic algorithm, namely fault detection, fault isolation, and fault identification [3, 19, 20]. The task of fault detection is to detect and report the abnormal conditions in the system being monitored. In fault isolation, the task is to locate the component which is failing or has been failed. In fault identification, the task is to estimate the nature of the fault when it is located [3, 20].

Prognostics on the other hand, deals with the prediction of failure before it occurs. The task of prognostic algorithm is to estimate the RUL or time to failure of the system [88]. In that sense, diagnostics is a posterior event analysis while prognostics is a prior event analysis [20]. The gap between the detection of the incipient faults and the progression to complete failure is the realm of prognostics [3]. In CBM, fault prognostics is superior to fault diagnostics because it can prevent the occurrence of failures, thus saves extra unscheduled maintenance cost. From the maintenance decision makers point of view, prognostics can be distinguished from diagnostics in terms of the warning time before failure (it is also called lead time or prognostic distance). In effective prognostic algorithms, such distance should be far enough ahead for an appropriate action to be taken [12].

Our focus in this doctoral research will be directly relevant to the prognostic algorithms in the CBM architecture. However, the performance of the prognostic algorithm is dependent on the performance of all the preceding procedures, in particular the diagnostics.

## 2.5 Fault prognostics in CBM

Based on the above discussion, the general scheme for the CBM program presented in this thesis, adopted from [89], is shown in Figure 2-2. The figure shows the entire offline and online procedures required for monitoring, diagnosis, and prognosis.

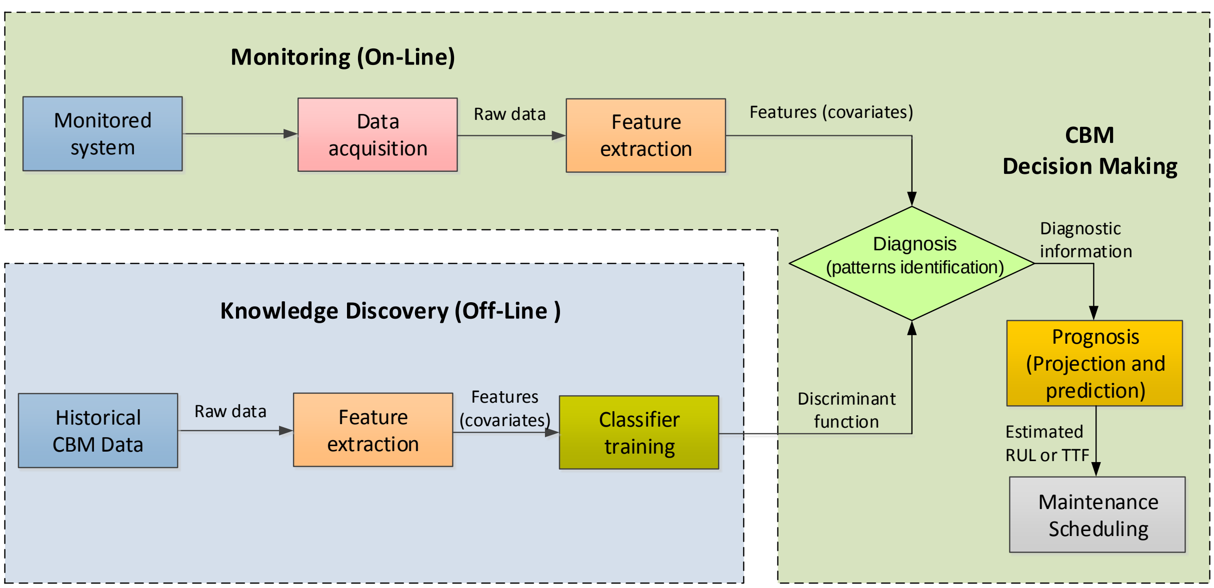


Figure 2-2: The overall architecture of the CBM program

As shown in Figure 2-2, two tasks should be implemented in advanced CBM decision making. The first one is the one that predicts the RUL of the monitored system, while the second task is the maintenance scheduling.

In the presence of a given fault, the function of the diagnostic model is to evaluate the current health condition of the monitored system, by performing a quantitative assessment between the newly arrived faulty signatures and the historical failure behaviors [22]. This assessment is very important for the prognostic algorithm to predict the RUL. The RUL estimation is then supplemented with forecasts describing the impact of predicted RUL on the operational and maintenance activities [30]. The RUL prediction is an essential task for the CBM decision maker, where a set of post-action decisions are taken. This is the task of maintenance scheduling module.

In maintenance scheduling, the objective is to identify the potential actions that could retard or eliminate the progression to the critical failure and prevent the future ones, based on the estimated RUL. This task takes into account some procedures such as logistics, inventory management, and maintenance planning [9]. These are management procedures, and they are out the scope of this doctoral research.

The accurracy of diagnostic algorithm is important to improve the prognostics, in the sense that the diagnostic information can be useful for updating some prognostic indices. One of the common prognostic indices in the CBM is the reliability function, which is used in this doctoral research to estimate the RUL of the monitored system.

## 2.6 Reliability-based prognosis

### 2.6.1 Reliability (survival) function

Reliability is defined in [12] as "*the ability of a product or system to perform as intended (i.e., without failure and within specified performance limits) for a specified time, in its life-cycle environment*". More specifically, it is the probability that a system will perform its intended function during a specified period of time, under certain operating or environmental conditions [28]. The reliability of a monitored system is affected by such conditions.

The object of primary interest in reliability analysis is the reliability function, conventionally denoted as $R(t)$. The reliability function $R(t)$ is also called the survival function or survivorship function, denoted as $S(t)$ in problems of survival analysis. In this thesis, the words 'reliability' and 'survival' are used interchangeably. The survival function is expressed mathematically as:

$$S(t) = Pr(T > t) = \int_t^{\infty} f(\tau) d\,\tau \qquad (2.1)$$

where $f(\tau)$ is the probability density function for the random variable $T$ which is denoting the system's failure time. Equation (2.1) states that the survival function is defined as the probability that the time of failure is later than some specified time $t$.

### 2.6.2 Survival function estimation: Parametric techniques

As stated previously, when the event data represent the failure mechanism, they are called lifetime data. Event data are helpful in reliability analysis. An example of event data are the failure times that are used to assess the survival function of a system. In parametric survival analysis, the event data are fitted to a predefined probability distribution function (for example, Weibull, exponential, gamma, etc.) to estimate the survival function. One of the limitations of such parametric techniques in survival analysis, is that some impractical statistical assumptions must be made about the distribution of the event data.

### 2.6.3 Non-Parametric survival curve: Kaplan-Meier estimator

Kaplan-Meier (KM) estimator is one of the most common non-parametric techniques applied to estimate the survival function, it is known as the product limit estimator [90]. The advantage of using KM estimator is that, it does not make any assumptions about the distribution of lifetime data.

The estimated KM survival function presented in [90, 91], is given as:

$$S(t) = \prod_{t_j \leq t} [1 - \frac{d_j}{n_j}] \tag{2.2}$$

where $n_j$ is the number of systems that have not failed (at risk) just before $t_j$, and $d_j$ is the number of failures at time $t_j$. More details about equation (2.2) are given in Appendix B of this thesis.

### 2.6.4 Prognosis: Remaining useful life estimation

The estimated survival function (either parametric or non-parametric), is used to predict the RUL of the monitored system, which in turn gives an indication to estimate its failure time [92]. This kind of prediction is called reliability-based prognosis.

The *Mean Remaining Useful Life* (*MRUL*) of a system at a certain time instant $t_k$, is calculated based on the following equation, which is presented in [93, 94]:

$$MRUL(t_k) = E(T - t_k | T > t_k) = \frac{\int_{t_k}^{\infty} (\tau - t_k) f(\tau) d\tau}{S(t_k)} = \frac{\int_{t_k}^{\infty} S(\tau) d\tau}{S(t_k)} \quad (2.3)$$

The interested readers may be referred to Appendix C for more details about this equation.

## 2.7 Updated survival curve using the condition monitoring data

### 2.7.1 The effects of the CM data on the survival curve

As stated previously, the event data are helpful in assessing the survival function and consequently the RUL of the system. However, the estimated survival function reflects only the effect of the age on the system's health state. It does not reflect the effect of the current operating conditions on its health state.

Using condition monitoring data in the CBM has several advantages. It is not necessary for the analyst to wait until the occurrence of complete failures. Instead, the analyst can use some condition indicators to predict the RUL of the system [92]. Thus, it is necessary to model the effects of condition monitoring data on the survival curve.

### 2.7.2 Proportional hazards model

Originally, the proportional hazards model (PHM) was applied in biomedicine to estimate the survival function for different groups of patients [95]. It is applied in the fields of reliability and CBM to estimate the reliability function of the monitored systems that work under different operating conditions [3].

The PHM is made up of two parts: the first part is the baseline hazard function, while the second part is a function including all the covariates that affect the time to failure of the system [13]. The parameters of the baseline hazard function can be estimated using many estimation techniques such as Breslow's estimator, while the effects of the covariates can be estimated using partial likelihood function [90]. The most common example for the baseline hazard function is the Weibull hazard function.

The principle of the PHM is discussed in Appendix D.

## 2.8 Multiple Failure Modes Prognostics in CBM

In the case of multiple failure modes prognostics, the objective is to estimate the RUL for a system subjected to a multitude of competing and interacting failure modes, by equating the system's RUL with the lowest failure time. In most statistical-based prognostic methods, the assumption of independence between the failure modes must be met.

It is concluded from the presented literature in [30] that the majority of multiple failure modes prognostic models do not help the industry practitioners to select an appropriate model for their specific needs, although these models have some technical merits. Other reported multiple failure modes prognostic models have been theoretical and restricted to a small number of failure modes, and do not concentrate on the practical implementation issues [96].

It is therefore necessary to model the survival function of a system or a component subjected to one of many failure modes, in the presence of the other competing ones.

## 2.9 Cumulative Incidence Function (CIF): A non-parametric technique for multiple failure modes

The cumulative incidence function (CIF) is a non-parametric model that does not need any statistical assumptions to be met about the distribution of the data, as introduced in [97]. In CIF estimation, the assumption of independence between the competing failure modes is not required to be considered. This technique uses the cumulative incidence rather than the survival probability [98]. It provides an estimate for the marginal probability of a certain failure mode in the presence of the other competing ones, based on the collected lifetime data. The CIF for a failure mode $i$, adopted from [97], is given as:

$$\hat{F}_i(t) = \sum_{\forall\, j, t_j \leq t} \frac{d_{ij}}{n_j} \hat{S}\left(t_{j-1}\right) \tag{2.4}$$

The CIF estimator for the failure mode $i$ depends not only on the number of systems that have experienced this type of failure ($d_{ij}$) at time $t_j$, but also on the number of systems that have not experienced any other failure mode ($n_j$) just before $t_j$. Equation (2.4) represents the probability that a system will experience a failure mode $i$ by time $t$, in the presence of the other competing failure modes.

In what follows, a literature review of the current prognostic models is discussed.

## 2.10 Current Prognostic Methods: A Literature Review

### 2.10.1 Diagnostic/Prognostic decision approaches

Numerous diagnostic and prognostic methods have been proposed and reported in the CBM literature [2, 3, 20]. A comprehensive literature review is found in [3], which is one of the most cited review papers in the CBM domain. According to that paper, current diagnostic/prognostic decision methods are broadly fall into two major categories: model-based and data-driven approaches. The model-based approaches are further classified into qualitative and quantitative methods [20]. The qualitative model represents the relationship between the inputs and outputs of a system in terms of qualitative functions centered on different units in the system. These kind of models are developed as qualitative causal models such as directed graphs (digraphs). Common examples of such models are Petri Nets (PNs) [99-101].

In quantitative model-based methods, an explicit dynamic mathematical model is developed based on the fundamental understanding of the system's physics [20]. Such models are capable of detecting and predicting the unanticipated faults, because they rely on residuals between the actual output of the system and its expected behavior (the output of the model) [2]. Such residuals represent the consistency checks between the sensed measurements of a real system and the outputs of a developed mathematical model [20]. The most common used techniques for generating and analyzing the residuals are: parameter estimation, parity relations, and observers [2, 3]. The physical models and damage propagation models which are based on damage mechanics, are employed commonly in the CBM to detect the incipient faults and to predict the failure as well [21]. Model-based approaches may not be practical in the case of complex systems that contain a multitude of components. The fault in such systems can vary from component to another.

In contrast to model-based approaches, data-driven approaches require transforming a sufficient amount of historical data into a priori knowledge, in order to build diagnostic/prognostic models [3]. They are based on the concept of training (also called learning) and testing, which keep improving as the knowledge is provided [102].

The data-driven approaches fall into two main categories: statistical-based and artificial intelligence-based methods. Artificial intelligence-based methods are the most popular and promising among the data-driven approaches [29]. Such methods rely on the availability of condition monitoring data and draw on learning techniques from the area of computational intelligence, where artificial neural networks (ANNs), Neuro-Fuzzy, and others are employed to map the condition measurements into fault growth models [61]. In the statistical-based methods, statistical machine learning algorithms are employed to process the acquired historical data, in order to discover the hidden patterns [43].

The advantage of utilizing the data-driven approaches over the model-based approaches in many practical cases is the ease to gather data rather than to build an accurate physical model [34]. The efficiency of data-driven approaches is highly dependent on the quantity and quality of acquired data [22]. Figure 3-1 is quoted from [20] to illustrate the difference between model-based and data-driven approaches.

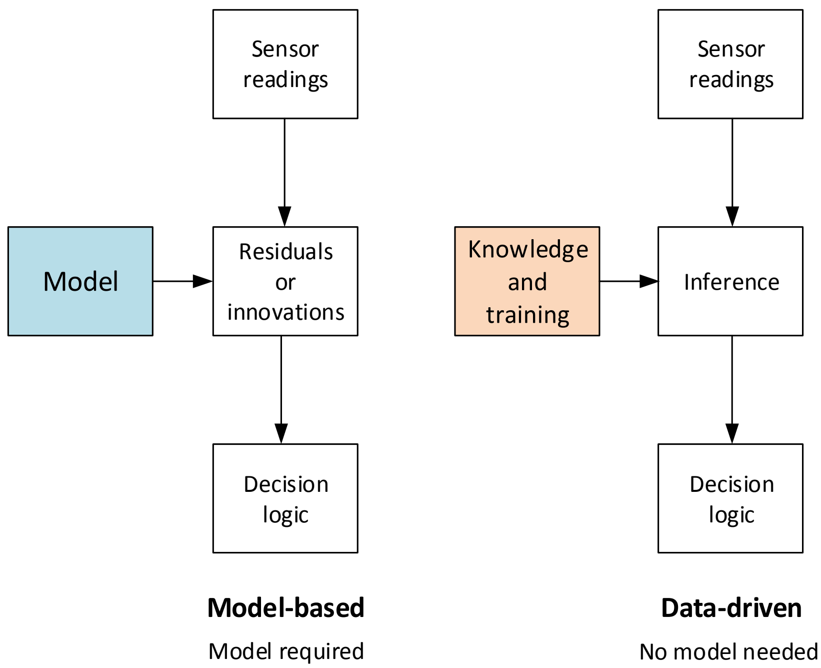


Figure 2-3: Model-based and data-driven approaches as presented in [20]

Since this doctoral research focuses mainly on prognostics, a literature review on the current prognostic methods will be considered in this thesis. A significant number of achievements in CBM prognostics are reviewed and reported in [30, 103-105]. Recently, numerous prognostic methods that have been implemented in many practical situations, are found in [30]. That paper presents a comprehensive survey on the recent prognostic methods used in the industry. It also discusses the theoritical and practical merits of each method. Other comprehensive literature reviews on prognostic models, have already been presented in [2, 3, 20]. The paper [106] surveys the prognostic methods that particularly applied to aerospace applications. Other good reviews on the rotating machinery prognostics are found in [42, 107, 108]. Those review papers present the state of the art in CBM prognostics.

The following sections present a literature review for the prognostic approaches employed in the context of CBM.

## 2.11 Model-Based Prognostic Approaches

A number of model-based methods have been successfully applied to real systems. The paper [109] presents a survey on the existing state of the art technologies of model-based fault diagnosis/prognosis for wheeled mobile robots (WMRs). The paper examines also a developed algorithm that is carried out on a WMR at the system and subsystem levels (power supply, driving, steering, suspensions, communication, and sensors). In that work a parametric estimation method is used to monitor the occurrence of some faults, such as flat tires deformation or broken spokes, while considering the parameters uncertainties. It is reported that the fault can be identified clearly using kinematic or dynamic models. The results in that paper conclude that model-based methods are used efficiently in such complex systems.

A physics-based model has been adapted and developed for the bearing prognostic applications in [110]. The Finite Element Analysis (FEA) is used to compute the spall growth trajectory and time to failure, based on given operating conditions. The method uses the bearing geometry, the load effect, and the speed of rolling element, to calculate the stress on the material surrounding the spall. Accordingly, the damage mechanics are used to model the spall propagation and to determine the cycles to failure. In that work, two different sets of seeded fault test data have been collected; one from an actual turbine engine bearing, while the other data were collected from smaller angular contact ball bearings. The bearings in the two tests were run under different loads and speeds.

A fault prognostic method based on both the physics of failure models and Bayesian estimation method was proposed in [60]. In that method, the particle filtering (PF) is used to provide a framework for long-term prognosis, while considering the uncertainties. The method is applied to an electrical power generation and storage (EPGS) system in automotive vehicles. The PF is used to predict the time evolution for the fault, based on typical automobile usage pattern and the temperature as a stress factor. The PF is used to estimate the probability density function by using a set of particles representing sampled values from the state space. In that work, the performance of the PF has been tested and compared with the extended Kalman filtering (EKF). The results showed that the prediction precision when using the PF is higher than that of using the EKF.

The battery is one of the most important component in our everyday life. One of the main prognostic issues when dealing with such kind of components is that the internal state variables are inaccessible to sensors [111]. As a model-based method, the PF was applied in battery health prognostics in [112]. In that work, the PF was compared to two data-driven statistical techniques namely; MM regression (it is the abbreviation of M in M estimates) and the robust-least trimmed sum of squares (LTS) regression, to predict the RUL of the battery. The battery health in that paper is directly tied to its capacity. The battery failed when its capacity has faded by 30%. In that paper, the RUL prediction using the PF approach got more accurate and precise prediction than the MM regression and the LTS regression.

A hybrid method that integrates the model-based and data-driven methods was proposed in [113]. The paper presents a systematic prognostic algorithm applied to automotive and electronic systems (the suspension system and battery). In that work, the hidden Markov model (HMM) that infers the evolution of component's degradation, is combined with the support vector regression (SVR) as a data-driven method. The proposed algorithm is designed from practical point of view, in order to allow the practitioners to carry out their experiments using an integrated framework.

As mentioned previously, it is not easy task to build an accurate physical model in most practical situations. Building an accurate model-based prognostic method may be a rather time-consuming and intensive process in the cases when the targeted systems are complex (e.g. aircraft turbine engines). Therefore, it is not advantageous to apply the model-based approaches in such cases.

## 2.12 Data-Driven Prognostics Approaches

### 2.12.1 Artificial Intelligence-based methods

As a data-driven prognostic method, the dynamic wavelet neural network (DWNN) was proposed in [114], to predict RUL of an industrial chiller. The rolling element bearing was used as the target component to demonstrate the applicability of the DWNN algorithm. Tri-axial accelerometers were employed to collect the vibration signals from a bearing with a crack in its inner race. The vibration signal was then preprocessed and a set of features were extracted. The input features to the DWNN are: the maximum values and the maximum power spectral density (PSD) of the measured signals in the three axial directions. The DWNN algorithm was implemented with seven neurons in the hidden node. The algorithm uses six features (the signal's amplitudes and the maximum PSDs) as inputs, and uses the width and length of the crack as outputs. The Mexican hats were used as mother wavelets functions. The trained DWNN model maps the features evolution to the crack growth. A failure threshold was established empirically to the crack width of 2000 μm or crack depth of 1000 μm, as an indication for the complete failure.

A prognostic method based on the ANNs was proposed in [115] to predict the failure time of the rolling bearings. The network was trained using the vibration-based degradation signals collected from a number of bearings. In that work, it was assumed that the bearing degradation signals possess an exponential growth. Each collected degradation signal was fitted to an exponential function and the parameters of that function were used to predict the failure time of the bearing. Another neural network called generalized regression neural network (GRNN) was used as an approximation technique to compute the regression model for each bearing. The regression model uses a weighted average of the exponential parameters coupled with parameter updating algorithm to predict the failure time of the bearing.

A prognostic method based on the recurrent neural networks (RNNs) have been conducted in [116]. In that work, a neural network prediction model called extended recurrent neural network (ERNN) was developed. The method deals with multiple condition monitoring indicators to predict the health conditions of a monitored equipment. The prediction performance using ERNN was tested and compared with the fully connected recurrent neural network (FCRNN) model, based on vibration data collected from a gearbox. The results show that the ERNN model gives better prediction results than the FCRNN model.

A prognostic method based on the ANNs has been proposed in [19] to predict the equipment's RUL. The method deals with multiple condition monitoring measurements. The inputs to the ANN model are the age and condition monitoring values at current and past time instants. The output is the life percentage. The life percentage in that work was defined as the ratio between the current age of the equipment and its failure time. The ANN model has four layers; an input layer, two hidden layers, and an output layer. A parametric function was proposed to fit the measured condition indicators. The optimal parameters' values of that function were obtained by using the genetic algorithm (GA). The fitted condition indicators were then used to train the ANN model instead of the original measurements, to avoid the representation of noise in the model. The trained ANN model represents the nonlinear mapping between the current and the previous condition indicators values, and the life percentage. The prognostic model was validated using vibration monitoring data collected from pump bearings.

In [117], a prognostic methodology based on a neural network was presented to predict the RUL of the aircraft actuator components. The methodology integrates both diagnostic and prognostic technologies. It exploits the diagnostic information throughout the operational life of the system to predict its RUL. It uses the neural network as signal processing and feature extraction technique, along with fuzzy logic classifiers and Bayesian inference as fusion strategies. The objective of the fusion module is to account for the diagnostic confidence. The system health state is implicitly modeled through the monitoring of specific features (such as the pressure, current, and position measurements). The methodology was applied to F/A-I8 stabilator electro-hydraulic servo valves (EHSVs). It is concluded from the results that the methodology can be implemented over a wide range of similar systems, including hydraulic and electro-mechanical actuation systems.

A reliability-based prognostic methodology incorporating failure and suspended data is proposed in [34]. The instantaneous reliability of a set historical items is first estimated using a variation of the KM estimator and a degradation-based failure probability density function. The estimated reliability is then used as the training target for a feed-forward neural network (FFNN), to predict the failure time five intervals ahead. The estimated survival probabilities are collected to form an estimated survival curve. A vibration signals collected from a set of pump bearings were used to validate the proposed methodology. The predicted failure time was identified by detecting the time at which the monitored unit has a survival probability less than a certain threshold value (a value of 0.5 was set in that work). As reported in that paper, the proposed prognostic methodology was able to capture the nonlinear relationship between the condition indicators and the actual health state of the monitored items.

A limitation of using the ANNs in prognostics is the difficulty of determining the optimal values of network parameters (number of nodes in the input layer, number of hidden layers, number of nodes in each hidden layer, learning rate, momentum, etc.). Finding the optimal number of hidden layers and the appropriate ANN structure remains a challenge and requires a number of experimentations. Another limitation of using these networks is their black box representations. The architecture of the network could not give a clear relationship between the inputs and outputs. As a consequence, there is no physical interpretation of the training process (the maintenance personnel could not have any conclusion about the relationship between the fault and its causes) [31]. Therefore, the ANNs should be merged with other transparent approaches such as the fuzzy logic approach, in order to have a clear view for the CBM decision making procedure.

In [118], similarity-based method was suggested to detect and predict the faulty behavior in a manufacturing plant. The historical data in the form of condition indicators (features) were collected from several identical machines. The degradation of two different machines were compared by assessing the similarity between any pair of feature vectors. In that work, a match matrix is calculated based on a similarity measure which is a function of Mahalanobis distance. The degradation patterns of the past and current operational run are compared through this match matrix. The future values of the match matrix indices are predicted using ARMA model or recurrent neural network (RNN). Given a set of signatures describing the faulty behavior, the prediction of the distribution of the feature vectors is used to estimate the future probabilities of failure. The method is computationally demanding since it compares the current operational run with a large number of past runs.

Another similarity-based prognostic method was presented in [119]. A set of failure trajectory patterns (reference trajectories) are extracted and collected in a reference library. A fuzzy-based similarity analysis is performed to predict the RUL of a newly developed failure trajectory pattern (test trajectory). The matching between the reference and test patterns is performed by fuzzy distance evaluation algorithm. The RUL estimation is performed in four steps [119]: (1) computing the fuzzy distance between the test and the reference patterns, (2) calculating their fuzzy similarity and distance score, (3) calculating the weights of the RUL estimates provided by the reference patterns, and (4) calculating the system's RUL by aggregating those weights. A nuclear power plant was employed as a case study to demonstrate the applicability of the proposed method.

A reliability-based prognostic methodology was proposed in [120] to update the reliability of a hazardous gas detection system (HGDS). The HGDS was used in aerospace shuttles to detect the explosion of gases. In that method, the failure data are assumed to follow an exponential distribution. In the first step of the proposed methodology, the system initial reliability parameters are calculated using Monte Carlo Simulation (MCS), by considering the variability of the system's conditions. Secondly, the fuzzy system is used to capture the effect of environmental factors on system's reliability. The input variables to the fuzzy system are: temperature, humidity, and the acidity. They were used as environmental factors. The output variable of the fuzzy system is a reliability estimate adjustor. In the last step, as the new failure data become available, the Bayesian inference uses the output of the fuzzy system to update the initial reliability estimated by the MCS. The proposed methodology thus allows continuous updating for the system reliability as new failure data are available at any stage of the system's life.

The limitation of applying the fuzzy logic in prognostics is the potential exponential explosion in the number of rules as the number of input variables increases [31]. Another limitation is its dependency on human expertise. Although this expertise is very valuable, it is subjected to some sort of tolerated performances, which often come from imprecise knowledge and inaccurate reasoning. For these reasons, those methods are often combined with some statistical tools and fusion algorithms to measure the uncertainty and confidence interval [20].

### 2.12.2 Statistical-based methods

Many statistical-based prognostic methods were utilized to predict the degradation conditions of the equipment [3, 20, 92]. In those methods, the model can be built from the observed degradation features using many statistical inference techniques. The PHM [13], logistic regression (LR) [121], and support vector machines (SVMs) [122], are the most common statistical inference procedures used in predicting the RUL in CBM.

In the fields of reliability and maintenance engineering, the PHM was employed to estimate and relates the multiple degradation indicators of the monitored equipment to specific reliability indices [123, 124].

The PHM was used in [125] to predict the raplacement time for the cutting tools in machining processes. It was used to model the tool's reliability and hazard function. The objective is to find the replacement time when the tool is used under different values of cutting speed, based on three criteria. In that paper, the PHM is proven to be a good model in representing tool reliability and hazard rate when the cutting speed is varying.

In [13], the PHM model was developed for the purpose of risk prediction. The objective is to identify the risk function which is merged with the cost data, in order to select the optimal maintenance policy. The model was applied to vibration signals collected from shear pump bearings in food processing plants, in three directions: axial, horizontal, and vertical. Such measurements were transformed by using the FFT into the frequency domain. Twenty-one covariates were used to build the PHM model. The significant covariates were then checked statistically and only three of them were selected. After the PHM has been built, the transition probabilities indicating the behavior of the covariates were estimated, to predict the occurrence of the failures. Accordingly, the optimal policy was followed to prevent their occurrence.

In the same context, the PHM was proposed to predict the hazard rate in [126]. The objective of the study is to maximize the reliability and to minimize the maintenance costs (multi-objective CBM optimization). At a certain inspection time, if the hazard rate multiplied by a predetermined scalar value $K$ is greater than a certain risk threshold value $d$, the item is replaced at this moment (preventive replacement). The operation continues if the value of the risk threshold could not be reached. In case of complete failure, a failure replacement is performed. The physical programming was applied as a multi-objective optimization technique, to obtain the optimal risk threshold value that optimizes the maintenance costs and reliability.

Unlike the PHM model proposed in [126] which mainly focuses on a single component, a PHM was proposed in [127] to deal with the CBM optimization in complex systems that consist of multiple components. In such systems, the economic dependency exists among the components. In that work, two levels of risk threshold (level-1 risk threshold $d_1$ and level-2 risk threshold $d_2$) were introduced, in order to determine which components should be subjected to preventive replacement, given that preventive replacements were performed for the other components at a certain inspection time.

The PHM was merged with the SVM to predict the RUL of a monitored equipment in [128]. The idea behind merging the SVM with PHM, is to use the SVM as a regression technique to detect the outliers (abnormal or extreme data). Then the trained SVM regression algorithm is used as a prediction model to calculate the relative loss in the data points. If the loss exceeds a certain value, the data point is considered to be abnormal. The dataset is then updated by eliminating the abnormal data points. The parameters of the PHM are then estimated based on the new updated training dataset. The PHM model is then used to predict the RUL of the monitored equipment. An experimental data were collected from oil samples in an engine to test the proposed prognostic method. The proposed method has a good short-time prediction capability as concluded in that paper. However, the long term prediction using that method was inaccurate. Choosing the optimal values of the SVM parameters is another issue in that method. Another algorithm is needed to optimize such values.

The logistic regression (LR) was widely used in statistical sciences and biomedicine [129]. The utilization of logistic regression in the CBM is still in its infancy [121]. The LR is a modified version of linear regression method that is employed when the dependent variable is a dichotomous variable (coded as 0 and 1) [43]. The idea behind this modification is to find the best fitting model that describes the relationship between a dichotomous variable and a set of independent variables (covariates). In the LR, a logistic function is used to represent the probability that an event will occur. It is constrained between 0 and 1 (logit transformations of the dichotomous variable) [122]. The resulting function (also called logit model) is a linear combination of the independent variables weighted by regression coefficients. After the dichotomous variable is transformed into a logit variable, the maximum likelihood estimation technique can be applied to calculate the parameters (the regression coefficients) of the model  [129].

A combination of the LR and relevance vector machine (RVM) was proposed in [121] as machine health prognostic methodology. In that work, the LR is used to estimate the failure degradation of the run-to-failure bearing data. The RVM is trained by using the estimated failure probability as the target output, and the kurtosis as an input feature. After the termination of the training phase, the RVM model is obtained in the form of weights and bias. The trained RVM model is then employed to predict the failure probability of the bearings. The trained model produced the output in terms of one-step-a head or *k*-step-ahead prediction [121]. The performance of the proposed methodology was validated using two datasets (simulated and experimental).

An LR prediction model was applied to estimate the RUL in condition monitoring in [130]. It was compared in that paper to the PHM prediction model. The comparison considers the efficiency and the computation efforts. The failure data and condition monitoring data were taken into account simultaneously in the two models. As a result of this comparison and from computational complexity point of view, the PHM considers the entire history of the degradation covariates. Thus a numerical integration procedure was carried out to estimate the model parameters. On the other hand, the LR model considers only the current degradation covariates. Consequently, the PHM required more computation efforts than that of the LR. The parameter estimation procedures in both models were carried out offline. This indicates that the computational complexity of each model is insignificant in the online applications. In that paper, an accelerated bearing test was conducted on a specially designed test rig. The experiment was conducted under elevated loads and rotational speed, and considered multiple degradation covariates.

Support vector machines (SVMs) originated from the statistical learning theory (SLT) [131-133]. More specifically, the SVM was introduced and developed by Vapnik and his co-workers [132, 134]. It is one of the most important classification techniques in the last 15 years due to its excellent generalization ability [135]. The SVM is known as maximum margin classifier which has the ability to minimize the empirical classification error and maximize the geometric margin between the training observations and the decision boundary simultaneously [133]. As a classification technique, it searches among the separating hyperplanes the one with the maximum margin by transforming the classification problem into an optimization problem. This optimization problem is treated as Lagrangian dual problem [136].

As reported in [132], for a smaller number of observations, the SVM has better generalization than the ANN when the local and global optimal solution are exactly the same. This can be considered as an advantage since the SVM can solve the same learning problem with a small number of observations.

SVMs were originally designed as two-class classifiers. Multi-class SVMs can be obtained by the combination of multiple two-class classifiers. Several methods have been proposed, such as one-against-one (OAO), one-against-all (OAA), and direct acyclic graph (DAG) [137]. In [138], a comparison between those three methods is presented. It is shown in that paper that the OAO method has outstanding performance than the other methods. The theory, methodology, and software of the SVM are available in [139]. The algorithms for SVM are also implemented in the publicly available software package "*Weka*" [43, 140].

In the context of CBM, the SVM has been applied in fault diagnosis and prognosis, due to its excellent performance and good generalization capability [22, 137]. A good survey on machine condition monitoring using the SVM is found in [137]. The paper [141] reviews the research and developments of the SVM in fault diagnosis and prognosis. That paper includes the application of the SVM to rolling element bearing, induction motors, machine tools, high voltage AC machines (HVAC), pumps, compressors, valves, and turbines.

The paper [141] applied the SVM as a regression technique called support vector regression (SVR) for machine fault prognosis. The SVR considers the predicted output to be numerical rather than categorical. A low methane compressor was presented as a case study to demonstrate the applicability of that method. The compressor was driven by 440 kW, 6600 volt, 2 poles motor. The operating speed of the motor was set to 3565 rpm. The data in that experiment were trending data of peak acceleration and envelop acceleration. The dataset contains observations of machine history with respect to the time. The objective of that method is to predict the future values of the vibration amplitude, based on those historical observations. The trained SVR learn the characteristics of the observations and save them in the form of weights, bias and support vectors, in order to predict the future conditions of the monitored machine.

A bearing prognostic methodology based on health state probability estimation was proposed in [142]. In that methodology, the fault diagnosis and health states estimation was performed using a range of classification algorithms, such as the ANNs, SVMs, regression trees, and others. The SVM classifier shows outstanding performance compared to the other classifiers. The RUL estimation was calculated using the classification ability of the SVM and probability distributions of the health states. Real life fault data collected from a set of bearings in high pressure-liquefied natural gas (HP-LNG) pump, were used to validate the feasibility of the proposed methodology. The optimal number of health degradation states was selected, and the results obtained in that work showed that the proposed prognostic methodology with five degradation states has the potential to be used for RUL prediction in the industrial applications.

Other statistical data-driven methods have been applied for the purpose of RUL prediction. Bayesian techniques are considered as good frameworks for the RUL prediction since they can handle various sources of uncertainties [143]. The techniques can perform dynamic state estimation, by constructing a probability density function, based on all available data [112]. Such techniques define the probability distributions over both parameters and covariates.

A prognostic method for wear prediction in oil-based monitoring was proposed in [144]. The method is based on stochastic filtering and continuous hidden Markov model (HMM). The Beta distribution is used to establish the transition matrix of the HMM, thus represents the unobservable state of the system in terms of wear. The relationship between the wear and metal concentration is modeled, based on observations representing the metal concentration at discrete time instants. The current and future states (the wear) of the system are predicted using recursive equations. In that paper, approximated grid and particle filtering were used as approximation techniques to compute such recursive equations.

The authors in [112] employed and compared statistical data-driven prognostic methods to predict the health state of the battery. The MM regression and robust-least trimmed sum of squares (robust-LTS) regression were explored as linear regression techniques, to extract the relationship between the charge transfer resistance $R_{CT}$ and electrolyte resistance $R_E$, and the capacity C at baseline temperature (25ºC). The feature that represents ($R_E$+$R_{CT}$) was extracted from the electrochemical impedance spectrometry (EIS) measurements at an elevated temperature (45 deg C). The Gaussian process regression (GPR) was employed in that work as a nonlinear probabilistic regression model to estimate the RUL.

The limitation of statistical-based methods is that they need sufficient amount of data to properly represent the nonlinear dynamics of the degradation process, in order to get more accurate and precise predictions. They may fail to learn the nonlinear trends in the absence of full range of training data. On the other hand, the parameter estimation procedures in such methods are complex and time consuming for historical data with a larger sample size. This will not facilitate the implementation of those methods in online applications. Moreover, those methods are based on statistical processes which require some impractical assumptions to be made.

### 2.12.3 Other approches: logical analysis of data

Logical Analysis of Data (LAD) is an interpretable knowlege discovery approach. It is based on the combinatorial optimization and Boolean functions [26]. It is proven in [26, 145] that the LAD approach can be used as a powerful diagnostic tool and reacts well to noisy and missing data.

The LAD approach was used in the field of CBM diagnosis for the first time, by the maintenance and reliability research team at École Polytechnique de Montréal, in 2007 [27]. The research team was able to reproduce human expertise in detecting the rogue components in airplanes [10].

The LAD approach generates patterns that can be easily interpreted and translated into rules which are beneficial to the maintenance engineers and the technicians [31]. The results presented in [31, 69] show that LAD is not based on any statistical analysis, this makes it capable of dealing with the condition monitoring data that are highly correlated.

The LAD approach has been used in [146] to predict the health states of an equipment, based on the patterns generated from the condition monitoring data.

In what follows, the stages of the LAD approach and its applications are described in more details.

## 2.13 LAD: Historical perespective

LAD is a supervised data mining approach that was presented in [26], by a group of researchers at RUTCOR, Rutgers University in USA. It was used for the first time as a Boolean classification technique to model the cause-effect relationship between a dependent variable that represents a certain event and a set of factors representing all the possible attributes that affect the occurrance of that event [147]. Originally, it is combinatorial optimization-based method used as two-class classification technique (*dichotomizer*) [26].

As reported in [26], the accuracy of LAD is superior to the other compared classification methods. In the field of CBM, it was also proved that the LAD approach is an efficient diagnostic tool when it is compared to the top classification techniques [10, 31].

### 2.13.1 Knowledge discovery in the form of extracted patterns

LAD is based on extracting the knowledge from a training dataset consisting of observations that are represented as binary or numerical vectors, contained in a number of classes. Each observation is composed of the values of certain characteristic factors [26]. In this thesis, the words ‘covariates’ and ‘factors’ are used interchangeably.

The knowledge is represented in the form of generated patterns for each class in the dataset. Those patterns represent the interactions between the factors in each class in the training dataset. The generated patterns are then used to construct a decision rule (model), that is used as a pattern-based classifier for the new testing observations that are not found in the training dataset [26].

In the conventional two-class LAD approach, the observations are classified as either positive $\Omega^+$or negative $\Omega^-$, where $\Omega^+$and $\Omega^-$ are the sets of positive and negative observations, respectively in the training data set $\Omega$ [31].

In what follows, we will discuss in details the stages of the LAD approach to generate the entire set of patterns that can be used to construct the decision model, for a single dichotomizer. Then we will explain how to generate the set of patterns to construct the multi-class LAD decision model.

## 2.14 Two-class LAD decision model

### 2.14.1 Stages of two-class LAD

The two-class LAD approach presented in [26] is composed of three stages: data binarization, pattern generation, and theory formation. In this thesis, the pattern generation stage is the main concern since the generated patterns represent the building blocks in the proposed CBM prognostic methodologies.

### 2.14.2 Data binarization

The data binarization stage involves the transformation of non-binary data, whether numerical, ordinal, or categorical to binary data, by transforming each non binary factor into a set of binary attributes.

The binarization of the non-binary data is a research subject that attracted many researchers. Many binarization techniques are available in the literature [24, 26, 145, 148, 149]. In this doctoral research, the binarization procedure presented in [26] is used as the first stage of the LAD approach. It is presented in details in Appendix E through a numerical example.

### 2.14.3 Pattern generation

Patterns generation is the key building block in the LAD approach. This step is essential in identifying the positive and negative patterns from the binarized dataset of positive and negative observations. The accuracy of the LAD decision rule depends on the type of generated patterns [150].

### 2.14.4 Definition and characteristics of patterns

A *literal* is either a binary attribute $b_j$ or its negation $\bar{b}_j$ [26]. A positive (negative) pattern is defined as an elementary conjunction of some literals that is true for at least one positive (negative) observation and false for all negative (positive) observations, in the training dataset. Each binary attribute in the training set can be represented in a pattern as a literal $b_j$ or its negation $\bar{b}_j$. The degree $d$ of a pattern indicates the number of literals used in its definition. In other words, a pattern $p$ of degree $d$ is a conjunction of $d$ literals. A pattern is said to cover a certain observation if it is true for that observation [26].

The set of observations covered by the pattern $p$ is denoted by $Cov(p)$. A high degree pattern is more likely to cover small proportion of observations, while the pattern with low degree is more likely to have large coverage [150].

By definition, the strictly defined patterns can cover some observations from one class but cannot cover any observations from the other classes [26]. This is called *pure* or *homogeneous* pattern. The concepts of pure patterns may be too restrictive in many practical situations since their coverage may be too limited [32]. The definition of the pattern is relaxed to allow the coverage of a large proportion of observations in one class, and a much smaller proportion of observations in the opposite class [23, 151]. This is called *non-pure* or *non-homogeneous* pattern. In other words, this relaxation allows the positive (negative) patterns to cover some negative (positive) observations. The homogeneity of the pattern is defined in [23].

Four types of patterns are defined in [26, 150, 152]: prime, spanned, strong and maximam-π pattern. A prime pattern is defined as the pattern that if any of its literals is eliminated, it will not be a pattern. The spanned pattern is composed of the possible maximum number of literals for the same number of covered observations [152]. A pattern $p_i$ is strong if there is no other pattern $p_k$ such that $Cov(p_i) \subset Cov(p_k)$. It is reported in [152] that the use of strong patterns leads to a superior performance of the LAD classifier. A maximal-π pattern is the one that has the largest coverage among all patterns that cover a certain observation $\pi$, in the training dataset. Two other types of patterns are defined as combinations of the prime, spanned, and strong patterns, namely strong prime and strong spanned patterns [150].

### 2.14.5 Pattern generation methods

In the literature, there are three common methods for pattern generation:

1- Enumeration-based methods [23, 26].

The enumeration-based methods are computationally demanding and time consuming for generating useful patterns if the dataset has a large number of binary attributes. In such case, the number of generated patterns in the training dataset can be extremely large. As an example, the number of degree $d$ patterns that can be generated from a binarized dataset of $n$ binary attributes is $2^d \binom{n}{d}$. For this reason, these approaches are limited to the datasets with small number of binary attributes.

2- Heuristic methods [151, 153].

In these methods, various heuristics have been developed and used for the solution of the pattern generation formulation problems. The patterns are generated from the training dataset by using heuristic algorithms applied iteratively until all observations are covered. These methods give feasible solutions but not the optimal ones. Moreover, their computational time is much shorter than the enumeration-based methods.

3- Mixed Integer and Linear Programming (MILP)-based methods [31, 150, 154].

The MILP-based method proposed in [150] can generate useful patterns that are optimal with respect to various selection preferences (simplicity, selectivity, and evidential [152]), without total enumeration. It can also generate patterns that also satisfy user specified requirements such as: the complexity (the degree of the pattern), coverage, and homogeneity [150]. The patterns generated by using the MILP-based method are strong patterns, which make the LAD classifier generalize better on the new testing observations.

### 2.14.6 MILP-based method

The basic task is the generation of positive and negative patterns from the binarized training dataset. In this thesis, we explain briefly the MILP-based method presented in [31]. The process of positive pattern generation is identical to that of negative pattern generation. For the sake of simplicity, the procedure for the generation of positive patterns is presented here, and stated in the following.

The procedure for generating one positive pattern $p$ is formulated into a set covering minimization problem whose decision variables are: the pattern vector $W$ of size $2n$ ($n$ is the number of binary attributes in the binarized dataset), the degree $d$, and the coverage vector $Y$ of size $|\Omega^+|$. The objective is to maximize (minimize) the number of observations that are (not) covered by the generated pattern $p$ in the set $\Omega^+$. The MILP formulation of the positive pattern generation procedure is summarized as follows:

$$\underset{W,Y,d}{minimize} \sum_{i \in \Omega^+} y_i$$

$$s.t. \begin{cases} w_j + w_{n+j} \leq 1 & \forall\, j = 1,2, \dots n & (2.5) \\ \sum_{j=1}^{2n} a_{i,j}\, w_j + n y_i \geq d & \forall\, i \in \Omega^+ & (2.6) \\ \sum_{j=1}^{2n} a_{i,j}\, w_j \leq d - 1 & \forall\, i \in \Omega^- & (2.7) \\ \sum_{j=1}^{2n} w_j = d & & (2.8) \\ 1 \leq d \leq n & & (2.9) \\ W \in \{0,1\}^n & & (2.10) \\ Y \in \{0,1\}^{|\Omega^+|} & & (2.11) \end{cases}$$

The feasible optimal solution $(W^*,\ Y^*, d^*)$ to the above MILP problem generates a pattern $p$ of degree $d^*$. The generated pattern is defined mathematically as:

$$p \coloneqq \bigwedge_{\substack{w_j = 1 \\ j \in \{1,..n\}}} b_j \bigwedge_{\substack{w_{j+n} = 1 \\ j \in \{1,..n\}}} \bar{b}_j \qquad (2.12)$$

More details about this MILP formulation are presented in Appendix F. A numerical example is given to clarify the procedures for generating the set of positive and negative patterns from a simple training dataset, as well.

It is proved in [150] that the generated patterns using the above MILP formulation are strong ones which are optimal with respect to the degree and coverage.

The above procedure for generating one positive pattern is repeated iteratively until every positive observation in the training dataset is covered by at least one generated pattern. If there are some uncovered observations, the above procedure is repeated after excluding the covered positive observations from the dataset. In other words, we remove the observations that are already covered by the generated pattern in the previous iteration, and the constraints (2.6) must be satisfied only for the uncovered positive observations. The procedures for generating the negative patterns are similar.

The MILP-based method proposed in [31] is a modified version of the method introduced in [150]. The modification aims at increasing the diversity of the generated patterns without a significant increase in the training time, thus increases the classification power of the two-class LAD decision model. The compact MILP-method proposed in [154] involves much smaller number of integer variables than the method presented in [150]. This participates in the reduction of computational burdens of the pattern generation stage in the LAD approach.

### 2.14.7 Pattern selection

In the pattern selection, a subset of the generated pattern is selected based on certain criteria [23, 26, 155]. The objective of this procedure is to find the minimal subset of patterns, such that each observation in the training dataset must be covered by at least one pattern [26]. In [23], the patterns are selected based on their prevalence, homogeneity, and degree. The pattern selection method proposed in [155], is based on solving a set covering problem (SCP). The method considers the coverage of the patterns as well as the outliers (the observations that are numerically far from the rest of the observations in the training dataset).

Like the pattern generation procedure, the selection procedure for the positive patterns is identical to the selection procedure for the negative ones. The SCP formulation presented in [155] is explained in the following.

We explain here the selection procedure for the set of positive patterns. Given the set of generated positive patterns deneoted by $P^+$, a Boolean vector $V = (v_1, v_2, \dots v_s, \dots v_{|P^+|})$ is assigned such that the binary variable $v_s$ is equal to one (zero) if the generated positive pattern $p_s$ is (not) selected in the required subset, where $|P^+|$ is the number of generated positive patterns. For the set of positive observations $\Omega^+ = \{a_1, a_2, \dots a_i, \dots\ a_{|\Omega^+|}\}$, a Boolean matrix $A$ of size $|\Omega^+| \times |P^+|$ is assigned such that its entry $A_{is}$ is defined as:

$$A_{is} = \begin{cases} 1 & \text{if the observation } a_i \text{ is covered by the pattern } p_s \\ 0 & \text{otherwise} \end{cases}$$

The SCP is formulated and given as follows:

$$minimize \sum_{s=1}^{|P^+|} v_s$$

$$s.t. \begin{cases} \sum_{s=1}^{|P^+|} A_{is}\, v_s \geq 1 & \forall\, i \in \Omega^+ \quad (2.13) \\ v_s \in \{0,1\} & \forall\, s \in P^+ \quad (2.14) \end{cases}$$

By solving this SCP formulation, a subset of positive patterns is selected and every positive observation in the training dataset is guaranteed to be covered by at least one selected pattern. The subset of negative patterns is selected in a similar way.

The above SCP is NP-complete [156]. Numerous heuristic approximation methods were well-studied and developed for its solution, for example reasonable and feasible solutions for the above SCP formulation can be easily obtained using the simple greedy algorithm, found in [155, 156]. Typically, such algorithm selects the patterns iteratively one-by-one until all the observations in the training dataset are covered. At each iteration, the algorithm selects the pattern that covers as many of the uncovered observations as possible.

### 2.14.8 Theory formation: discriminant function

The final stage in the LAD approach is the theory formation. For the conventional two-class LAD classifier, the positive and negative patterns are used to create a model called the discriminant function that generates a score ranging between -1 and 1. The discriminant function for the new unseen observation $O$ is constructed and given as:

$$\hat{\Delta}(O) = \sum_{i=1}^{N^+} W_i^+ p_i^+(O) - \sum_{i=1}^{N^-} W_i^- p_i^-(O) \quad (2.15)$$

where $W_i^+ (W_i^-)$ is the weight of the positive (negative) pattern $p_i^+(p_i^-)$. The value $p_i^+(O)(p_i^-(O))$ is equal to 1 if pattern $p_i^+(p_i^-)$ covers the observation $O$, and zero otherwise, and $N^+(N^-)$ is the number of generated positive (negative) patterns. The weight of the positive (negative) pattern is defined as the ratio between the number of covered positive (negative) observations by that pattern and the total coverage of the positive (negative) patterns. Accordingly, $W_i^+$ is defined as:

$$W_i^+ = \frac{cov(p_i^+)}{\sum_{i=1}^{N^+} cov(p_i^+)} \tag{2.16}$$

and $W_i^-$ is defined as:

$$W_i^- = \frac{cov(p_i^-)}{\sum_{i=1}^{N^-} cov(p_i^-)} \tag{2.17}$$

The output of the discriminant function has a positive value when the tested observation belongs to the positive class and a negative value otherwise. A value of zero means no classification is possible (unclassified observation) [31, 150]. The tested observation can be correctly classified, misclassified, or unclassified. The misclassified observations are the result of generating low degree patterns while unclassified observations are results of high degree patterns [150, 154].

The accuracy of the LAD decision model depends on the type of generated patterns. As experimented and reported in [150, 154], it is shown that the strong patterns can achieve a significant and superior testing accuracy. The experimentations in those papers show that the generated strong prime patterns can reduce the number of unclassified observations while the strong spanned patterns can reduce the number misclassified observations.

## 2.15 Multi-class LAD decision model

Like the conventional two-class LAD, the multi-class LAD approach presented in this thesis is composed of three steps: data binarization, pattern generation, and theory formation. In the following subsections, the pattern generation and the resulting decision model for the multi-class LAD approach, are discussed.

### 2.15.1 Pattern generation for multi-class LAD decision model

The generation of positive and negative patterns in two-class LAD model are extended in multi-class LAD decision model. In the literature, there are two approaches on the extension of the two-class LAD to multi-class applications [157, 158]. Those works present the different philosophies that are followed to break down the multi-class classification problems (*Polychotomy*) into two-class problems (*Dichotomies*). Another philosophy to build a multi-class LAD decision model was proposed in [69]. It involves modifying the architecture of the two-class LAD in order to construct a unified multi-class LAD classifier. It uses OAO method to built the multi-class LAD model. The proposed method has the advantage that it generates a less complex decision model which has a better execution time [69].

We discuss briefly the procedure presented in [69] to build a set of multi-class patterns that are used to create the multi-class LAD decision model. For the multi-class classification problem with $C$ classes, the procedure starts by creating a set of patterns $P_{ie}$ for each pair of classes $(c_i, c_e)$, where $i, e \in \{1,2,..C\}$ and $i \neq e$. In that method the set $P_{ie}$ is not identical to the set $P_{ei}$ (i.e. $P_{ie} \neq P_{ei}$). The paterns in the set $P_{ie}$ are generated through the solution of the MILP-based pattern generation formulation presented in [69]. Accordingly, $C(C-1)$ sets of patterns are constructed. The pseudocode for that procedure is presented in [69].

### 2.15.2 Scoring function for multi-class LAD decision model

The discriminant function used in the multi-class LAD approach is significantly different from that of the two-class LAD. A multi-class LAD decision model was proposed in [159], by using OAA method. In that decision model, $C$ sets of patterns are constructed. The pattern set $P_i$ ($i \in \{1,2,..C\}$) separates class $i$ from all the remaining $(C-1)$ classes.

The discriminant function presented in that work generates a score for each class and therefore the testing observation belongs to the class with the highest score. The score is calculated for a certain class, by using all the sets of generated pattern. For a new testing observation $O$, the score is calculated as given in [159] as:

$$\hat{\Delta}(O) = \underset{i=1,\ldots C}{arg\ max} \sum_{p_t^i \in P_i} p_t^i(O) w_t^i \tag{2.18}$$

where $p_t^i(O) = 1$ if the pattern $p_t^i$ covers the observation $O$ and zero otherwise. The value $w_t^i$ associated with the pattern $p_t$ in the set $P_i$ acts as a normalized weight. The output of equation (2.18) is the class with the highest score.

## 2.16 LAD's applications

### 2.16.1 Application of LAD in medical diagnosis and prognosis

LAD was used successfully in the medical field to diagnose patient’s condition and to predict the propagation of some diseases [23-25].

The risk of occurrance of a certain event (risk-of-death) in medical prognosis was estimated for two groups of patients in [23]. The dataset has two classes of observations representing the groups of patients who died and who survived during a follow-up period of 9 years. In that work, a modified discriminant function called prognostic index was defined for all patients. The value of the prognostic index was shown to be closely correlated with the patients' risk of death, and outperforms the widely used indicator by cardiologists; Cox Score [23].

A methodology called the Logical Analysis of Survival Data (LASD) was proposed in [25]. The objective of that methodology is to identify the interactions between the factors that affect the death of a group of patients. Each generated pattern covers only a proportion of patients in the dataset. A survival function for each pattern is estimated using the KM estimator. The patterns' survival curves are used to construct a survival model for the new patient that is not included in the training dataset. Each of those survival curves are participating with the same weight in that survival model [25]. The performance of the LASD was compared to the survival decision trees and the KM estimator. The empirical results show that the LASD is an accurate prognostic model and outperforms the compared techniques.

As an important conclusion from the above two medical prognostic applications, the high degree of correlation among the factors can be detected and used to estimate the survival curves, without making any statistical assumptions.

### 2.16.2 Other applications

The LAD approach was applied in the airlines industry to estimate the overbooking level by predicting the show rates of the passengers [160]. The idea of that paper is to extract the set of patterns that cover passengers who have higher or lower show rates from a given training dataset consisting of a number of observations. Each observation represents a passenger and is characterized by a set of attributes. According to the extracted patterns, the passengers are classified as one of the two outcomes show or no-show.

The LAD approach was applied in the field of finance and banking as a credit risk rating model [161]. In that paper, a dataset consisting of a number of banks is used to extract the patterns that are characterizing the banks having high ratings and those having low ratings. The discriminant function is constructed using the extracted patterns to evaluate the credit quality of banks. It is shown from the presented results that the LAD credit risk model is an accurate predictive tool in comparison to the other compared statistical models; the ordered logistic regression and the support vector machines.

### 2.16.3 Applications of LAD in CBM diagnostics

The LAD approach was used in the field of CBM for the first time by the maintenance and reliability research team at École Polytechnique de Montréal, Canada [27]. In that work, the first version of the software *cbmLAD* [162] was developed for the CBM diagnostic applications.

In 2009, the research team was able to detect and analyze the phenomenon of rogue components in airplanes [10]. In that research, LAD was explored for the purpose of detecting rogues within a population of repairable components. The LAD approach was compared to the most popular techniques used in the industry and its performance was comparably better than those techniques.

The application of LAD in the detection of the faulty bearings, is presented in [31]. It was used in that work as a tool for automatic diagnosis of the faults in rolling bearings, by using a modified MILP-based pattern generation method. In that paper, LAD was compared to the SVM, neural networks, and others and it is shown that it outperformed those techniques.

The diagnosis of faults in power transformers using the multi-class LAD, was proposed in the doctoral research of Mortada [32]. The objective of that research was to design a tool for detecting and identifying the transformers' faults, by extracting the patterns from the dissolved gas analysis (DGA) data. An updated version of the software *cbmLAD* was issued as a result of that doctoral research.

## 2.17 Application of LAD in CBM Prognostics: Proposed methodologies

The advantages mentioned in the introduction of this thesis motivate us to apply the LAD approach in CBM prognostics. In this doctoral research, LAD is applied in three novel prognostic methodologies. The philosophy behined each methodology is to exploit the constituents of the historical CBM data (the time and covariates) properly, in order to extract the hidden knowledge. Consequently, if this knowledge is extracted in an adequate manner, this can lead to an accurate RUL prediction for the monitored system. This philosophy is followed in each methodology, by combining LAD as an event driven technique with a time-driven estimation technique.

In each of these methodologies, based on the lifetime data, the survival function of the system is estimated by using a non-parametric reliability estimation method. Such method does not make any assumption for the distribution of the lifetime data. The estimated survival function represents the hidden knowledge in the first constituent of the CBM data (the time). It describes the temporal characteristics of the historical systems. LAD on the other hand is exploited to extract the knowledge from the covariates (the second constituent of the CBM data). It handles the condition monitoring data, in order to extract useful patterns that reflect the effect of the operating conditions on the survival function of the monitored system.

In the first and second methodologies, the objective is to predict the RUL of a monitored system working under different operating conditions, while considering the analysis of single failure mode. The third methodology deals with the RUL prediction of a system working under different operating conditions, in the presence of multiple failure modes.

The two methodologies for single failure mode prognostics differ in the way of representing the data. In both methodologies, the two-class LAD is merged with the KM as a non-parametric estimator.

In the first methodology, the prognostic knowledge is extracted from the historical CBM data that are collected from a set of systems at/or immediately before the occurrance of failure. In this case, the data consist of the aging times and corrersponding covariates. In the second methodology, the knowledge is extracted from all the lifespan data that are collected from the historical systems. For each system, the data consist of all observations collected during its operational period until the occurrance of failure. More specifically, each observation contains the inspection times and corrersponding covariates.

In the third methodology, all the lifespan data collected from a set of historical systems subjected to multiple failure modes, are exploited to extract the prognostic knowledge. The multi-class LAD is merged with a set of cumulative incidence functions as non-parametric estimators.

Detailed discussions of these methodologies and their respective steps, together with validation procedures and tests to accomplish the stated research objectives, are presented in the subsequent chapters.

# CHAPTER 3 ARTICLE 1 : REMAINING USEFUL LIFE PREDICTION USING PROGNOSTIC METHODOLOGY BASED ON LOGICAL ANALYSIS OF DATA AND KAPLAN–MEIER ESTIMATION

Ahmed Ragab, Mohamed-Salah Ouali, Soumaya Yacout, Hany Osman

Published in: JOURNAL OF INTELLIGENT MANUFACTURING

DOI 10.1007/s10845-014-0926-3

## 3.1 Abstract

Most of the reported prognostic techniques use a small number of condition indicators and/or use a thresholding strategies in order to predict the remaing useful life (RUL). In this paper, we propose a reliability-based prognostic methodology that uses condition monitoring (CM) data which can deal with any number of condition indicators, without selecting the most significant ones, as many methods propose. Morever, it does not depend on any thresholding strategies provided by the maintenance experts to separate normal and abnormal values of condition indicators. The proposed prognostic methodology uses both the age and condition monitoring data as inputs to estimate the RUL. The key idea behind this methodology is that, it uses Kaplan-Meier (KM) as a time-driven estimation technique, and Logical Analysis of Data (LAD) as an event-driven diagnostic technique to reflect the effect of the operating conditions on the age of the monitored equipment. The performance of the estimated RUL is measured in terms of the difference between the predicted and the actual RUL of the monitored equipment. A comparison between the proposed methodology and one of the common RUL prediction technique; Cox proportional hazard model, is given in this paper. A common dataset in the field of prognostics is employed to evaluate the proposed methodology.

## 3.2 Introduction

One way to minimize both maintenance costs, and probability of failure is to perform an assessment and prediction of asset health and future failures based on current health, operating conditions, and maintenance history (Kothamasu et al. 2006). Condition based maintenance (CBM) utilizes condition monitoring technologies in order to detect and predict the future health states of the engineering assets, based on non-intrusive measurements of their current health conditions (Tan et al. 2009).

The acquired data in a CBM system can be categorized into two main categories (Jardine et al. 2006): condition monitoring data and event data. Condition monitoring data are collected and processed to determine the equipment health condition. These measured observations are related to the health states of the monitored equipment (Jardine et al. 2006). Also, they reflect the degradation conditions of such equipment (Hamada 2005). Event data provide the information on what happened (failure, installation, overhaul, etc.) and/or what was done (repair, preventive maintenance, oil change, etc.) to the targeted equipment (Jardine et al. 2006). When the event data represent the failure mechanism, they are called lifetime data. An example of event data is the failure times that can be used to develop the survival (also called reliability) function of the equipment. In the parametric survival analysis, the event data are fitted to a known probability distribution (Weibull, exponential, gamma distributions, etc.) (Elsayed 2012). Non-parametric techniques are used to estimate the survival function without making any statistical assumption about the probability distribution (Klein and Moeschberger 1997). One of the most common non-parametric techniques is Kaplan–Meier (KM) estimator (Klein and Moeschberger 1997). The estimated survival function using either parametric or non-parametric techniques, is used to predict the failure time of the monitored equipment which in turn allows to predict its remaining useful life (RUL) (it is called reliability-based prognosis) (Jardine et al. 2006). The RUL (also called residual life) is defined as the time left before the occurrence of complete failure (Vachtsevanos et al. 2006). Using condition monitoring data in CBM system has several advantages. It is not necessary for the analyst to wait until the occurrence of complete failures. Instead, the analyst can use some condition indicators to predict the RUL of the equipment.

In CBM prognostic, a particular issue is to estimate the RUL of the monitored equipment under different operating conditions. This is because there is no clear and understandable relationship between the condition measurement and the prediction of RUL. Prognosis requires precise and adaptive models to estimate future equipment health states and to predict its RUL (Tian et al. 2010). Fortunately, the advanced sensors and signal processing techniques enable the maintenance practitioners to extract multiple degradation features for the purpose of degradation detection (Kim et al. 2012). Recently, RUL prediction has received an increasing attention as industrial equipment become complex and critical (e.g. aircraft turbine engine) (Schwabacher and Goebel 2007). In many industrial situations, prognostic of equipment performance relies on human experience. Although, the experts may have significant experience about equipment failure and degradation states, they do not have systematic methodology that can predict the RUL of the monitored equipment (Kothamasu et al. 2006). Therefore, there is a need to develop and improve prognostic schemes that can be implemented in CBM with minimum human involvement.

Numerous diagnostic and prognostic models have been proposed in the CBM literature. They broadly fall into two major categories: model-based approaches and data-driven approaches. In model-based approaches, explicit dynamic models based on the fundamental understanding of the physics of the system, are developed to detect significant faults, and to predict catastrophic failures (Vachtsevanos et al. 2006). These approaches may not be the most practical approach because there is no easy way to obtain an accurate model. In contrast to model-based approaches, data-driven approaches require transforming a sufficient amount of historical data into a prior knowledge to build behavior models. They are based on the concept of training and testing, which keeps improving as the knowledge is provided. One advantage of utilizing data-driven approaches in many practical cases is the availability of advanced sensing technologies that collect data, rather than to build an accurate physical model (Heng et al. 2009).

The data-driven prognostic approaches fall into two main categories: statistical approaches, and artificial intelligent approaches. Artificial intelligent approaches are the most popular and promising data-driven approaches (Schwabacher and Goebel 2007). These approaches rely on the availability of condition monitoring data and draw on learning techniques from the area of computational intelligence, where artificial neural networks (ANNs), and other techniques are employed to map measurements into fault growth models (Vachtsevanos et al. 2006; Duda et al. 2001).

A data-driven prognostic method based on artificial neural networks (ANNs) has been proposed in Tian et al. (2010). That method aims to predict the RUL, while dealing with condition monitoring measurements. A reliability based prognostic method based on ANNs is proposed in Heng et al. (2009). In that method, the instantaneous reliability of items is calculated using a variation of Kaplan–Meier estimator and a degradation-based failure function. One of the limitation of that prognostic method is that it uses only one condition indicator to estimate that degradation-based function. Another limitation of that method is the consideration of assigning a certain threshold to the survival curve in order to predict the failure time. In general, a common drawback of using ANNs in prognostics is the difficulty of determining the optimal values of the network parameters (i.e. number of nodes in the input layer, the number of hidden layers, the number of the nodes in each hidden layer, the learning rate, etc.). Another drawback of using these networks is the black box representation, the architecture of the network could not be characterized.

In the statistical data-driven prognostics, the statistical machine learning algorithms are employed to process the acquired historical data in order to discover the hidden patterns in such data (Witten et al. 2011). In the context of CBM prognostic, many statistical data-driven approaches were utilized to predict the degradation conditions of the equipment (Jardine et al. 2006; Vachtsevanos et al. 2006). In those approaches, the model can be built from the observed degradation features using various statistical inference techniques. Proportional hazards model (PHM) (Jardine et al. 1999), logistic regression (LR) (Caesarendra et al. 2010), and support vector machines (SVMs) (Friedman et al. 2001) are the most common statistical inference techniques used for the RUL estimation. PHM is employed in the field of the CBM to estimate the reliability function. It relates the multiple degradation indicators to specific reliability indices of the monitored equipment (Jardine et al. 2006). Also, PHM was used in Tian et al. (2012) to predict the risk of failure of equipment working under different operating conditions. LR model is used along with the relative vector machine (RVM) in Liao et al. (2006) to estimate the failure degradation of run-to-failure bearing data. SVM is used for fault diagnosis and prognosis due to its excellent performance and good generalization capability (Kim et al. 2012; Widodo and Yang 2007). Other statistical data-driven approaches have been applied for RUL estimation such as Bayesian technique which is considered as a good framework since it can handle various sources of uncertainties (Saha and Goebel 2008). Hidden Markov model (HMM) is proposed in Wang (2007) as a prognostic method for wear prediction in oil-based monitoring. A statistical prognostic method is proposed in Son et al. (2013), it combines Wiener process as stochastic process, and principal component analysis, in order to model the degradation in a set of historical systems, and to estimate the RUL of another set of similar systems. Statistical prognostic methods possess some drawbacks since they are based on statistical inference procedures that require some impractical assumptions to be met.

In the field of CBM, a combinatorial optimization-based method called logical analysis of data (LAD) was used for automatic diagnosis of faults in rolling bearings in Mortada et al. (2011). The results in Bennane and Yacout (2012) indicate that LAD is a promising tool in CBM diagnosis and prognosis. LAD was used on a great variety of classification problems and reacts well to data noise and measurement errors (Bores et al. 2000). Since LAD is not based on any statistical assumptions, it can deal with several covariates that may be highly correlated. Another important advantage of LAD in CBM diagnostic is the transparency of the technique, which leads to clear interpretability of diagnostic results (Mortada et al. 2011). LAD relies on extracting patterns from a set of training observations (Bores et al. 2000). The patterns can be easily interpreted and translated into rules which can be used by maintenance engineers and technicians. LAD can deal with any type of data: event data, condition monitoring data, or both (Mortada et al. 2011).

According to the previously mentioned advantages, applying LAD approach in CBM prognostic should be an interesting challenge. The challenge, though, is how to develop prognostic models using all condition monitoring data collected under different operating conditions, without making any statistical assumption, or assigning any prior thresholds. The aim of this paper is to propose a new prognostic methodology based on LAD in order to predict the RUL of a set of equipment working under different operating conditions.

This paper is divided into five major sections: Section 3.3 presents LAD along with the definition of patterns, the pattern generation approaches, and the idea of combining LAD to KM estimator. Section 3.4 introduces the proposed prognostic methodology and explains its procedures in details through a numerical example. Section 3.5 presents a common case study in the field of CBM prognostics to validate the proposed methodology. It also presents a comparison between the results obtained by using our proposed methodology, and those obtained by using Cox proportional hazard model (Cox PHM). Section 3.6 concludes the paper and proposes some future works to be followed.

## 3.3 Logical Analysis Of Data

### 3.3.1 Logical analysis of data: preliminaries

Logical analysis of data is a supervised data mining, pattern generation and classification technique that was introduced in (Bores et al. 2000). LAD was used as a Boolean technique to identify the causes of a certain event through investigating a set of factors representing all the possible causes of that event (Crama et al. 1988). It was applied on a various classification problems and the results in Bores et al. (2000) indicate that LAD's classification accuracy is competitive and often superior to the other classification methods.

LAD is used to extract knowledge from a dataset consisting of observations that can be represented as binary or numerical vectors. Each observation is composed of the values of certain covariates. Originally, LAD was used as two-class classification technique that is a dichotomizer (Bores et al. 2000). As a combinatorial and optimization method, it has been evolved as an effective decision model that relies on extracting patterns from binarized data in order to formulate decision rules that classify data into two classes; called positive and negative (Mortada et al. 2011). Each extracted pattern represents the interactions between the covariates of either positive or negative observations in the training dataset. Accordingly, LAD can be used as pattern-based classifier for the new observations that are not included in the training dataset (Bores et al. 2000).

The two-class LAD approach is composed of three stages: data binarization, pattern generation, and theory formation. Pattern generation is the key building block in LAD decision model. This stage is essential in identifying a set of positive and negative patterns from the training dataset of positive and negative observations. In this paper, the main concern is the pattern generation stage since the set of generated patterns constitute one of the corner stones of the proposed prognostic methodology. The following subsection presents the definition and characteristics of the patterns, and reviews some of the pattern generation techniques. For more details about LAD, the interested readers may be referred to Bores et al. (2000), Crama et al. (1988), Mortada et al. (2013) and Alexe et al. (2007).

### 3.3.2 Pattern generation for two class LAD decision model

A positive (negative) pattern is defined as an elementary conjunction of literals that is true for at least one positive (negative) observation and false for all negative (positive) observations in the training dataset (Kim et al. 2012). The set of observations covered by the pattern $p$ is denoted as $Cov(p)$. The strictly defined patterns, which are called pure patterns, cover some observations from one class but do not cover any observations from the opposite class (Bores et al. 2000). The non-pure patterns cover a large proportion of the observations in one class, and a much smaller proportion of the observations in the opposite class (Mortada et al. 2013; Alexe et al. 2003). The accuracy of LAD decision model depends on some characteristics of the generated patterns (Alexe et al. 2007; Hammer et al. 2004).

In the literature, there are three common approaches for pattern generation: enumeration-based approaches (Bores et al. 2000), heuristic approaches (Hammer and Bonates 2006), and mixed integer linear programming (MILP)-based approaches (Mortada et al. 2011; Ryoo and Jang 2009). In MILP based approaches, the objective is to maximize the number of positive (negative) observations that are covered by the generated positive (negative) patterns, while generating patterns that are optimal with respect to certain preferences or constraints (Ryoo and Jang 2009).

The MILP based method proposed in Guo and Ryoo (2012) involves much smaller number of 0–1 integer variables than the approach presented in Ryoo and Jang (2009), thus requires shorter training time. The MILP based method proposed in Mortada et al. (2011) is a modified version of the approach introduced in Ryoo and Jang (2009). It aims at maximizing the diversity of the generated patterns from the training dataset without a significant increase in training time, thus increases the classification power. For more details about the pattern generation techniques, the interested readers may be referred to Mortada et al. (2011), Bores et al. (2000), Hammer et al. (2004), Ryoo and Jang (2009) and Guo and Ryoo (2012).

### 3.3.3 Combining LAD to KM: the survival curves of the generated patterns

In this paper, a prognostic methodology based on LAD and KM is proposed in order to predict the RUL of a set of equipment working under different operating conditions. The main idea is to use Kaplan-Meier as a time-driven estimation technique and LAD as an event-driven diagnostic technique. The baseline survival curve, called Kaplan-Meier curve, reflects only the effect of the age on the health state of the monitored equipment. LAD is used to generate a set of patterns that represent the interactions among the covariates without making any prior statistical assumption. A set of survival curves (one curve for each generated pattern) are estimated using Kaplan-Meier estimation model. The observations collected from the monitored equipment, are then employed to update its baseline survival curve according to the patterns covering those observations. This updating is carried out by averaging the weighted sum of the survival curves of the covering patterns and the baseline survival curve. Based on the updated survival curve, the RUL is estimated given that the equipment has survived up to the current time instant, and the patterns covering the recent observation.

The main idea of our proposed methodology is inspired from the logical analysis of survival data (LASD) proposed in (Kronek and Reddy 2008), that was applied in the field of medical prognosis. In that paper, the survival curve for a given observation is updated by averaging the baseline curve and the survival curves of the patterns that cover that observation. Each of the survival curves of the patterns has a weight equal to that of the baseline, although it does not cover the same number of observations covered by the baseline, in the training dataset. This is one of the limitations of the updating formula in that methodology.

In this paper, we propose two modified versions of the formula proposed in (Kronek and Reddy 2008) to update the survival curve of the recent observation. In the two proposed formulas, the survival curve of the monitored equipment is updated initially by averaging the survival curve of the patterns that cover the first updating observation and the baseline survival curve. Then, the following updating observations are used to update the survival curve by averaging the survival curves of the patterns that cover each observation and the former updated survival curve. We also go further in this paper by providing a single measure to evaluate the RUL based on the area under the updated survival curve.

More specifically, in the first proposed updating formula, the survival curve for each of the generated patterns has the same weight as that of the baseline survival curve. In the second proposed formula, we modify the first formula by considering a weight for each of the survival curves of the patterns that is less than the weight of the baseline curve. In other words, the weight of the baseline curve is greater than the weight of the survival curve of each pattern. The next section presents the methodology in details, through a number of steps.

## 3.4 Proposed Prognostic Methodology

Based on a set of equipment monitoring data, the proposed prognostic methodology exploits the capability of Kaplan-Meier estimator and the two-class LAD. The objective is to predict the RUL of the monitored equipment based on its own operating conditions and condition indicators. Unlike the baseline survival curve (Kaplan-Meier curve) which is based on all the failure observations, we decompose these failure observations into two different categories (classes); failure observations that characterize the *Short Life* (*SL*) equipment and call it the positive class, and the other failure observations that characterize the *Long Life* (*LL*) equipment, which we call negative class. The equipment that fails before a certain time $t_S$ specified and decided by the maintenance personnel is called *SL* equipment, and the one that fails after that time is called *LL* equipment. A feasible and practical choice for the time $t_S$ is the *mean time to failure* (*MTTF*). Our objective is to use the most recent available observations (updating data) about the operating conditions and the condition indicators in order to give better estimation of the RUL for both *SL* and *LL* equipment. Figure 3-1 depicts the schematic diagram for the proposed prognostic methodology.

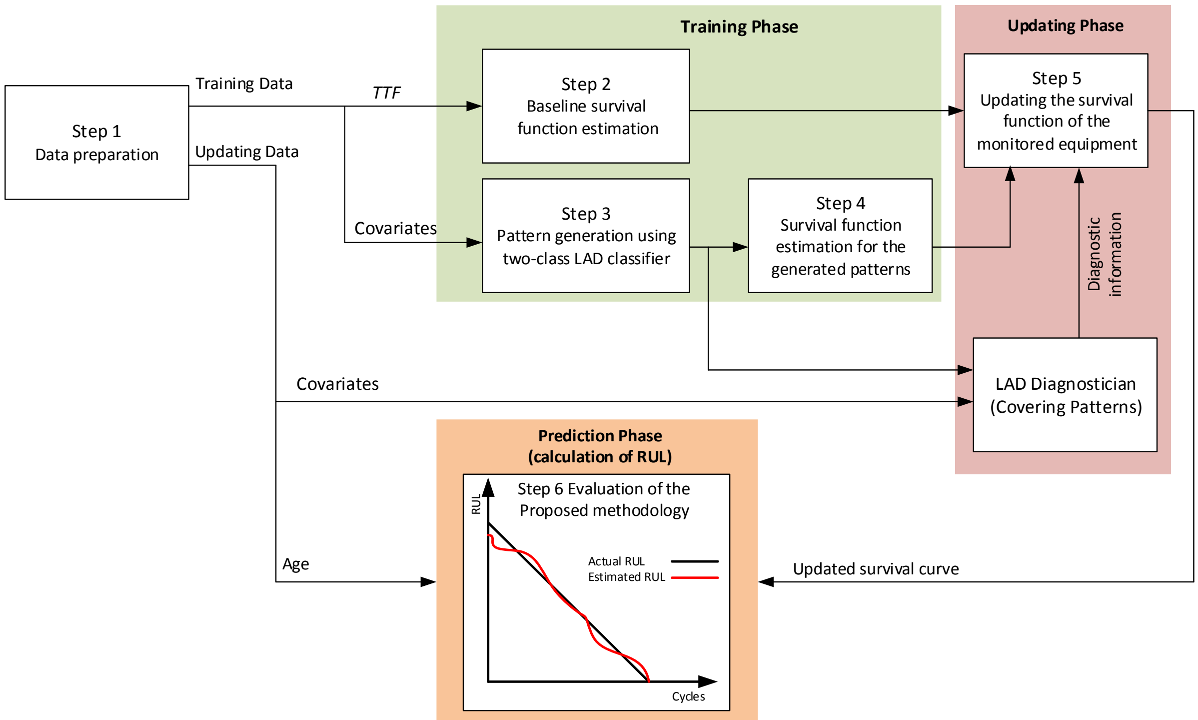


Figure 3-1: Schematic diagram for the proposed prognostic methodology

The stated objective of the methodology is achieved through the following steps:

**Step 1**: **Data preparation**. In this step, the data is prepared and partitioned into two subsets; training dataset and updating dataset. The training data contains a set of historical failure observations for the two classes of equipment; *short life* and *long life* equipment. The updating dataset consists of a set of observations collected from another set of similar equipment.

**Step 2**: **Baseline survival curve estimation**. In this step, the baseline survival curve is estimated using *Kaplan-Meier* estimation technique for all equipment, by using only the lifetime data in the training dataset prepared in step 1.

**Step 3: Pattern generation using two-class LAD method.** In this step, the patterns are generated using two-class LAD methodology for the two classes of equipment, the *SL* and the *LL*, by using only the covariates that are representing the operating conditions and the condition indicators in the training dataset prepared in step 1.

**Step 4: Survival curve estimation for the generated patterns**. In this step, the survival curve for each pattern generated in step 3, whether *SL* or *LL* pattern, is estimated using Kaplan-Meier estimation technique by considering only the set of equipment covered by this pattern.

**Step 5: Updating the survival curve of the monitored equipment**. Given a set of recent observations from the updating dataset which are collected from similar type of equipment, the survival curve of each equipment is updated using one of two different updating formulas. This updating step exploits the data indicating the operating conditions and the recent condition indicators in order to give an early indication of whether the equipment seems to be an *SL* or an *LL*. It is used to give a better prediction of the RUL.

**Step 6: RUL estimation**. The RUL of the monitored equipment is estimated by using the updated survival curve obtained in step 5. The estimated RUL is predicted and compared to the actual RUL at each time instant.

In order to illustrate the proposed methodology, we use a fictitious numerical example of ten equipment working under five covariates. The numerical example is designed in a way that facilitates the introduction of the methodology to the reader. The validation of the proposed methodology is given in section 4 where we use a common dataset in the realm of prognostics; the Turbofan engine dataset (C-MAPSS dataset) which is available on the website of NASA prognostic data repository (Saxena, et al. 2008). The dataset consists of observations collected from a set of equipment working under different operating conditions. Each of the six previously mentioned steps is explained in details in the following subsections.

**Step 1: Data preparation for training, and updating**

The training and updating datasets are prepared from a set of collected observations. Each observation comprises the identity of the equipment, the time or age, and a set of covariates reflecting the operating conditions and the condition indicators as well. Each observation in the training dataset consists of the *time to failure (TTF)* of the equipment and several covariates (i.e. operating conditions and condition indicators). The updating dataset consists of a group of observations collected from another set of similar equipment. Each observation in this dataset consists of the time (age) of the equipment and several covariates.

**Training dataset.** Table 3.1 presents the training dataset for $T$ equipment, where each equipment has only one failure observation consisting of the time to failure and a number of covariates. The equipment are classified according to their *TTF* as either $\Omega^{SL}$or $\Omega^{LL}$, where $\Omega^{SL}$and $\Omega^{LL}$ are the sets of $SL$ and $LL$ equipment, respectively in the training dataset $\Omega$. The time to failure of each equipment in the training dataset is used to estimate the baseline survival curve as stated previously in step 2 in the methodology. The covariates are used by LAD to generate the patterns that are discriminating between the set of $SL$ equipment and the set of $LL$ equipment, as stated in step 3 of the methodology. In Table 3.1, $O(i, t_{Fi}, Z_{i,t_{Fi}})$ represents the failure observation collected from the $i^{th}$ equipment ($i = 1,2, \ldots T$), $Z_{i,t_{Fi}}$ is the corresponding vector of covariates, $t_S$ is the separation time between the two classes ($SL$ and $LL$), and $t_{Fi}$ is the failure time of the $i^{th}$ equipment.

Table 3.1: Representation of the training dataset for $T$ equipment

| | | |
|---|---|---|
| Training dataset $\Omega$ | $SL$ equipment $\Omega^{SL}$ (positive observations) | $O\left(i, t_{Fi} \leq t_S, Z_{i,t_{Fi} \leq t_S}\right)$ |
| | $LL$ equipment $\Omega^{LL}$ (negative observations) | $O\left(i, t_{Fi} > t_S, Z_{i,t_{Fi} > t_S}\right)$ |

**Numerical Example**

In this example, we consider two classes of equipment ($SL$ and $LL$). Each class contains a group of equipment that failed at different time units. The equipment are working under different operating conditions. The time that separates the two classes is set to be nine time units ($t_S = 9$). The first class contains the observations at failure of the five equipment that failed at 6, 8, 4, 5, and 7 time units, respectively, while the second class contains the observations at failure of the five equipment that failed at 10, 11, 12, 13 and 14 time units, respectively. Table 4.2 shows the $SL$ and the $LL$ observations with the failure times and five covariates' measurements, $Z_{i,t_{Fi}}$, for the $i^{\text{th}}$ equipment, where $i = 1,2, \ldots, 10$. These covariates represent the operating conditions and the condition indicators. The operating conditions are controllable, for example the speed and the mechanical load. In real life, they are set to certain values by the user. The condition indicators are gathered as sensory information reflecting the equipment's conditions, for example the temperature, the vibration, and the percentage of oil debris. These conditions change with the time. In Table 3.2, we note that the working conditions may have different values, since LAD can deal perfectly with this situation. Also in Table 3.2, we assume that the covariates are already transformed from numerical to binarized values, by using the binarization stage of LAD.

Table 3.2: Dataset for *SL* and *LL* equipment

| Equipment Identity | Class | Time To Failure (TTF) | Covariates | | | | |
|---|---|---|---|---|---|---|---|
| | | | $Z_1$ | $Z_2$ | $Z_3$ | $Z_4$ | $Z_5$ |
| 1 | *SL* (positive) | 6 | 1 | 0 | 1 | 1 | 1 |
| 2 | | 8 | 0 | 0 | 0 | 1 | 1 |
| 3 | | 4 | 1 | 1 | 1 | 1 | 1 |
| 4 | | 5 | 1 | 1 | 1 | 0 | 1 |
| 5 | | 7 | 1 | 1 | 1 | 0 | 0 |
| 6 | *LL* (negative) | 10 | 1 | 0 | 0 | 1 | 0 |
| 7 | | 11 | 0 | 0 | 1 | 0 | 1 |
| 8 | | 12 | 1 | 0 | 1 | 0 | 0 |
| 9 | | 13 | 1 | 0 | 0 | 0 | 0 |
| 10 | | 14 | 0 | 0 | 1 | 0 | 0 |

**Updating dataset.** We utilize the set of observations collected from another set of similar equipment at every one operational cycle (the number of equipment in the updating dataset is $U$). This dataset contains both failure and functional observations. The operating conditions, the condition indicators, and the cycle time in each observation are included. The updating observation $O(u, t_k, Z_{u,t_k})$ represents the observation collected from the $u^{th}$ equipment ($u = 1,2, \ldots U$) at time $t_k$ ($t_k = 1,2, \ldots t_{Fu}$), and $Z_{u,t_k}$ is a vector of covariates representing the operating conditions and condition indicators at time $t_k$.

In this numerical example, we have a number of updating observations collected from another equipment (let us call it equipment 11). The updating observations for that equipment is presented in Table 3.3. The observations are different in age as well as the values of the operating conditions and the condition indicators. The equipment fails after the 8$^{th}$ cycle. All observations in the second column are functional observations except the last one shown in gray which is a failure observation (the 8$^{th}$ observation is the last observation collected before failure).

Table 3.3: Updating observations from equipment 11

| Equipment identity | Observation | $t_k$ | $z_1$ | $z_2$ | $z_3$ | $z_4$ | $z_5$ |
|---|---|---|---|---|---|---|---|
| 11 | $O(11, 1, Z_{11,1})$ | 1 | 1 | 0 | 0 | 0 | 0 |
| | $O(11, 2, Z_{11,2})$ | 2 | 1 | 1 | 1 | 0 | 0 |
| | $O(11, 3, Z_{11,3})$ | 3 | 0 | 0 | 0 | 1 | 1 |
| | $O(11, 4, Z_{11,4})$ | 4 | 0 | 1 | 1 | 1 | 0 |
| | $O(11, 5, Z_{11,5})$ | 5 | 1 | 0 | 1 | 1 | 1 |
| | $O(11, 6, Z_{11,6})$ | 6 | 0 | 0 | 0 | 1 | 1 |
| | $O(11, 7, Z_{11,7})$ | 7 | 1 | 1 | 1 | 0 | 0 |
| | $O(11, 8, Z_{11,8})$ | 8 | 0 | 0 | 0 | 1 | 1 |

**Step 2: Baseline survival curve estimation using Kaplan-Meier (KM)**

The baseline survival curve (also called survival probability function), is estimated using KM estimation technique (Klein and Moeschberger). The inputs to KM estimator are the failure times of all the *SL* and *LL* equipment. The baseline curve is estimated using equation (3.1) as follows:

$$S_b(t) = \prod_{t_{Fi} \leq t} [1 - \frac{d_{t_{Fi}}}{Y_{t_{Fi}}}] \tag{3.1}$$

where $d_{t_{Fi}}$ is the number of equipment that failed at time $t_{Fi}$, and $Y_{t_{Fi}}$ is the number of equipment which are at risk at time $t_{Fi}$. By "at risk", we mean the number of equipment that did not fail before $t_{Fi}$. The baseline $S_b(t)$ considers all the failure observations of the two classes (*SL* and *LL* equipment). Note that $S_b(t)$ does not reflect the effect of the operating conditions on the health state of the equipment. It reflects only the effect of the working time (or age) on its health state.

From the third column in Table 3.2, the baseline survival function $S_b(t)$ is calculated using the *TTFs* of the ten equipment. It is shown in Table 3.4 and plotted as a stepwise curve in black color in Figure 3-2. The highlighted cells in Table 3.4 represent the values of the baseline survival function at the time instants that are not shown in the third column of Table 3.2.

Table 3.4: The baseline survival curve

| Time | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_b(t)$ | 1 | 1 | 1 | 1 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0 |

**Step 3: Pattern generation using two-class LAD method**

In this step, we use LAD to generate the *SL* and *LL* patterns that differentiate between the *SL* and the *LL* equipment in the training dataset. A specific characteristic of LAD is the extraction of collection of patterns which are the interactions between the operating conditions and the condition indicators for *SL* and *LL* equipment in the training dataset. As mentioned previously in Section 3.3, a pure pattern can cover one or more observations from one class and non of the observations in the opposite class. Based on the training data in Table 4.1, two-class LAD diagnostic is used to generate the *SL* and *LL* patterns. Each of the generated patterns may cover only a subset of observations in the training dataset, but each observation must be covered by at least one pattern. LAD keeps searching for patterns until all the observations in the training dataset are covered.

In this numerical examples, LAD found five patterns that cover all the observations in the dataset in Table 3.2. From this dataset, the five covariates and the class labels (*SL* and *LL*) are used to generate the *SL* and *LL* patterns. Table 3.5 shows the *SL* and *LL* patterns that are generated by LAD for this example and the corresponding covered equipment. The three *SL* patterns and two *LL* patterns guarantee that each of the *SL* and *LL* observations in Table 3.2, is covered by at least one pattern. The five generated patterns listed in the table are pure patterns.

Table 3.5: Generated *SL* and *LL* patterns

| The generated *SL* Patterns | Interpretation | Covered equipment *Cov(p)* |
|---|---|---|
| $p_1$ | $Z_1 Z_5$ | 1, 3, 4 |
| $p_2$ | $Z_4 Z_5$ | 1, 2, 3 |
| $p_3$ | $Z_2$ | 3, 4, 5 |
| **The generated *LL* Patterns** | **Interpretation** | **Covered equipment *Cov(p)*** |
| $p_4$ | $\overline{Z_2}\ \overline{Z_5}$ | 6, 8, 9, 10 |
| $p_5$ | $\overline{Z_2}\ \overline{Z_4}$ | 7, 8, 9, 10 |

**Step 4: Survival curve estimation for the generated patterns using KM estimation**

The survival curve of each generated pattern is calculated by considering only the equipment covered by that pattern. Accordingly, the inputs to KM estimation technique are the failure times for the equipment that are covered by that pattern. The survival curve of the *j*th pattern ($j = 1,2, \dots |P|$, where $P$ is the set of generated patterns) is estimated using equation (3.2) which is a modified version of equation (3.1), as follows:

$$S_{p_j}(t) = \prod_{\substack{t_{Fi} \le t \\ \forall\, O(i, t_{Fi}, Z_{i,t_{Fi}}) \in cov(p_j)}} \left[1 - \frac{d_{t_{Fi}}}{Y_{t_{Fi}}}\right] \tag{3.2}$$

It is noticed from Table 3.5 that pattern $p_1$covers three *SL* equipment (equipment 1, 3, and 4). Its survival probability will be 2/3 after the failure of equipment 1. The estimated survival curve of the pattern $p_1$ is depicted in the third row of Table 3.6. This curve is plotted in Figure 3-2 in red. Similarly, the estimated survival curves of the patterns $p_2$, $p_3$, $p_4$, and $p_5$ are depicted in Table 3.6, and are plotted in Figure 3-2.

Table 3.6: Survival curves of *SL* patterns, *LL* patterns, and the baseline

| Time | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_b(t)$ | 1 | 1 | 1 | 1 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0 |
| $S_{p_1}(t)$ | 1 | 1 | 1 | 1 | 0.667 | 0.333 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $S_{p_2}(t)$ | 1 | 1 | 1 | 1 | 0.667 | 0.667 | 0.333 | 0.333 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $S_{p_3}(t)$ | 1 | 1 | 1 | 1 | 0.667 | 0.333 | 0.333 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $S_{p_4}(t)$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.75 | 0.75 | 0.5 | 0.25 | 0 |
| $S_{p_5}(t)$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.75 | 0.5 | 0.25 | 0 |

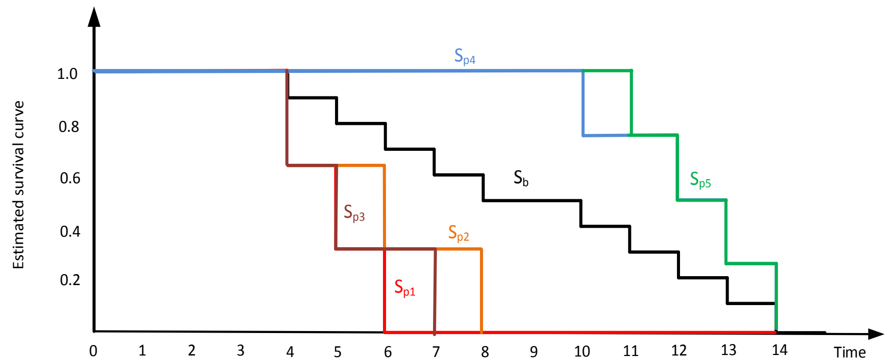


Figure 3-2: Estimated survival curves for the baseline, and the five generated patterns

**Step 5: Updating the survival curve of a monitored equipment**

The estimated survival curves (the baseline survival curve and the survival curves for the *SL* and *LL* patterns) are used to update the survival curve of a certain monitored equipment. The observations collected from the equipment are employed in order to extract the hidden patterns and to use these patterns to update its survival curve. The updated survival curve reflects the age of the equipment, its operating conditions, and its condtions indicators' values. Initially, the survival curve after the first updating observation is calculated by averaging the survival curve of the patterns that cover that observation and the baseline survival curve. The following updating observations are then used to update the survival curve by averaging the survival curves of the patterns that cover each observation and the former updated survival curve simultanuously. One of the following two models, is used separately to update the survival curve of the monitored equipment $u$. The two models are represented by formula 1 and formula 2 respectively, as follows:

**Formula 1**

$$S_{O(u,t_k,Z_{u,t_k})}(t) = \begin{cases} [\sum_{j=1}^{n} S_{p_j}(t) + S_b(t)\,]/(n+1) & \forall\, k = 1 \\ [\sum_{j=1}^{n} S_{p_j}(t) + S_f(t)\,]/(n\,+1) & \forall\, k \geq 2 \end{cases} \tag{3.3}$$

**Formula 2**

$$S_{O(u,t_k,Z_{u,t_k})}(t) = \begin{cases} [\sum_{j=1}^{n} S_{p_j}(t)/n\, + S_b(t)\,]/2 & \forall\, k = 1 \\ [\sum_{j=1}^{n} S_{p_j}(t)/n\, + S_f(t)\,]/2 & \forall\, k \geq 2 \end{cases} \tag{3.4}$$

where $n$ is the number of patterns that cover the updating observation, $S_b(t)$ is the baseline curve, and $S_f(t)$ is the former updated survival curve obtained from the previous updating observation at time $t_{k-1}$. In formula 1, the survival curve is updated by assigning the same weight for each of the survival curves of the covering patterns and the baseline. In formula 2, the survival curve is updated by assigning a weight for the baseline curve greater than the weight of the survival curve of each pattern covering the observation.

To illustrate the updating procedure clearly, we use the updating observations listed in Table 3.3 to update the survival curve of equipment 11, by using formula 2. Two-class LAD is used as a diagnostic technique to find the patterns that cover each updating observation simultaneously. Each updating observation is covered by a set of patterns as shown in the second column of Table 3.7. Each row in Table 3.7 represents the updated survival curve after collecting each updating observation.

Table 3.7: Updating procedure for the estimated survival curve (formula 2)

| **Observation/ Time** | **Covering Patterns** | **Updated survival** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $O(11,1,Z_{11,1})$ | $p_4, p_5$ | $S_{O(11,1,Z_{11,1})}(t)$ | 1 | 1 | 1 | 0.95 | 0.9 | 0.85 | 0.8 | 0.75 | 0.75 | 0.638 | 0.525 | 0.35 | 0.175 | 0 |
| $O(11,2,Z_{11,2})$ | $p_3$ | $S_{O(11,2,Z_{11,2})}(t)$ | 1 | 1 | 1 | 0.808 | 0.617 | 0.592 | 0.4 | 0.375 | 0.375 | 0.319 | 0.263 | 0.175 | 0.088 | 0 |
| $O(11,3,Z_{11,3})$ | $p_2$ | $S_{O(11,3,Z_{11,3})}(t)$ | 1 | 1 | 1 | 0.738 | 0.642 | 0.463 | 0.367 | 0.188 | 0.188 | 0.159 | 0.131 | 0.088 | 0.044 | 0 |
| $O(11,4,Z_{11,4})$ | $p_3$ | $S_{O(11,4,Z_{11,4})}(t)$ | 1 | 1 | 1 | 0.702 | 0.488 | 0.398 | 0.183 | 0.094 | 0.094 | 0.08 | 0.066 | 0.044 | 0.022 | 0 |
| $O(11,5,Z_{11,5})$ | $p_1, p_2$ | $S_{O(11,5,Z_{11,5})}(t)$ | 1 | 1 | 1 | 0.684 | 0.494 | 0.282 | 0.175 | 0.047 | 0.047 | 0.04 | 0.033 | 0.022 | 0.011 | 0 |
| $O(11,6,Z_{11,6})$ | $p_2$ | $S_{O(11,6,Z_{11,6})}(t)$ | 1 | 1 | 1 | 0.676 | 0.580 | 0.308 | 0.254 | 0.023 | 0.023 | 0.02 | 0.016 | 0.011 | 0.005 | 0 |
| $O(11,7,Z_{11,7})$ | $p_3$ | $S_{O(11,7,Z_{11,7})}(t)$ | 1 | 1 | 1 | 0.671 | 0.457 | 0.321 | 0.127 | 0.012 | 0.012 | 0.01 | 0.008 | 0.005 | 0.003 | 0 |
| $O(11,8,Z_{11,8})$ | $p_2$ | $S_{O(11,8,Z_{11,8})}(t)$ | 1 | 1 | 1 | 0.669 | 0.562 | 0.327 | 0.230 | 0.006 | 0.006 | 0.005 | 0.004 | 0.003 | 0.001 | 0 |

$Z_{11,1} = (1,0,0,0,0)$, $Z_{11,2} = (1,1,1,0,0)$, $Z_{11,3} = (0,0,0,1,1)$, $Z_{11,4} = (0,1,1,1,0)$, $Z_{11,5} = (1,0,1,1,1)$, $Z_{11,6} = (0,0,0,1,1)$, $Z_{11,7} = (1,1,1,0,0)$, $Z_{11,8} = (0,0,0,1,1)$

To explain how the values in Table 4.7 are calculated, we consider, for example, the 6th cycle (the highlighted column). The survival curve of the first updating observation $O(11,1,Z_{11,1})$ is calculated by averaging the baseline survival curve and the survival curves of the patterns that were found by LAD to cover that observation (the $LL$ patterns $p_4$ and $p_5$). Accordingly, the survival curve of equipment 11 is estimated initially after collecting the first updating observation as follows:

$$S_{O(11,1,Z_{11,1})}(t) = [\sum_{j=1}^{n} S_{p_j}(t)/n + S_b(t)]/2 \quad = [(S_{p_4}(t) + S_{p_5}(t))/2 + S_b(t)]/2$$

From Table 4.6, the baseline has the value of 0.7 and the values of the survival curves for the patterns $p_4$ and $p_5$, are 1 (one) and 1 in the 6th cycle, respectively (see the 8th column in Table 3.6).

Accordingly, the value of the updated survival curve at the 6th cycle after collecting the first updating observation, is calculated as follows:

$$S_{O(11,1,Z_{11,1})}(6) = [\sum_{j=1}^{n} S_{p_j}(6)/n + S_b(6)]/2 \quad = [(S_{p_4}(6) + S_{p_5}(6))/2 + S_b(6)]/2$$

$$= [(1+1)/2 + 0.7]/2 = 0.85$$

The second updating observation (the observation $O(11, 2, Z_{11,2})$ is found to be covered by the *SL* pattern $p_3$. The survival curve is updated by averaging the survival curve of the pattern $p_3$ and the former updated survival curve as:

$$S_{O(11,2,Z_{11,2})}(t) = [\sum_{j=1}^{n} S_{p_j}(t)/n + S_f(t)]/2 \quad = [S_{p_3}(t)/1 + S_f(t)]/2$$

The value of the updated survival curve at the 6th cycle after collecting the second updating observation, is calculated as follows:

$$S_{O(11,2,Z_{11,2})}(6) = [S_{p_3}(6)/1 + S_f(6)]/2 = [0.333/1 + 0.85]/2 = 0.592$$

Similarly, the survival curve at each time of collecting a new updating observations, is calculated. This updating procedure guarantees that the resulting survival curve will be monotonically decreasing.

**Step 6: RUL estimation**

It is important to represent the effect of the operating conditions and the condition indicators in order to accurately predict the RUL of the monitored equipment. From the updating observations, we have updated the survival curve that reflects not only the age but also the operating conditions of the equipment. The remaining useful life of the monitored equipment is predicted after each time its survival curve is updated. Let $T$ represents the time to failure which is a random variable and let $T - t_k$ represents the RUL of $T$ at time $t_k$. The *mean remaining useful life* (*MRUL*) is calculated using equation (3.5) (Banjevic and Jardine 2007) as:

$$MRUL(t_k) = \frac{\int_{t_k}^{\infty} S(\tau)d\tau}{S(t_k)} \tag{3.5}$$

where $S(\tau)$ is the survival function. In this prognostic methodology, the *MRUL* is calculated at each time instant when new updating observation is collected, by considering the updated survival curve. The updated survival curve $S_{O(u,t_k,Z_{u,t_k})}(t)$ of equipment $u$ is used instead of $S(\tau)$ in equation (5), and the *MRUL* for that equipment is calculated as:

$$MRUL_u(t_k) = \frac{\int_{t_k}^{\infty} S_{O(u,t_k,Z_{u,t_k})}(\tau)d\tau}{S_{O(u,t_k,Z_{u,t_k})}(t_k)} \tag{3.6}$$

The calculation of the *MRUL* is performed through dividing by the value of the survival probability one period before, since the survival curve is estimated at discrete instants of time (Pintilie 2006). Accordingly, the value $S_{O(u,t_k,Z_{u,t_k})}(t_{k-1})$ is used in the denominator of equation (3.6) instead of $S_{O(u,t_k,Z_{u,t_k})}(t_k)$. Experimentally, equations (3.6) can be represented in discrete form as follows:

$$MRUL_u(t_k) = \frac{\sum_{t_r=t_k}^{\infty} \Delta t_r S_{O(u,t_k,Z_{u,t_k})}(t_r)}{S_{O(u,t_k,Z_{u,t_k})}(t_{k-1})} \tag{3.7}$$

where $\Delta t_r$ is the monitoring (inspection) interval defined as: $t_r = t_{r+1} - t_r$.

Based on the updated survival curves of equipment 11, calculated in Table 3.7, the values of MRUL are calculated using equation (3.7). These values are listed in Table 3.8, and are ploted in Figure 3-3.

Table 3.8: MRUL calculation for equipment 11

| Time: $t_k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Actual RUL | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| $MRUL_{11}(t_k)$ | 9.69 | 6.01 | 4.01 | 2.17 | 1.68 | 1.14 | 0.55 | 0.11 |

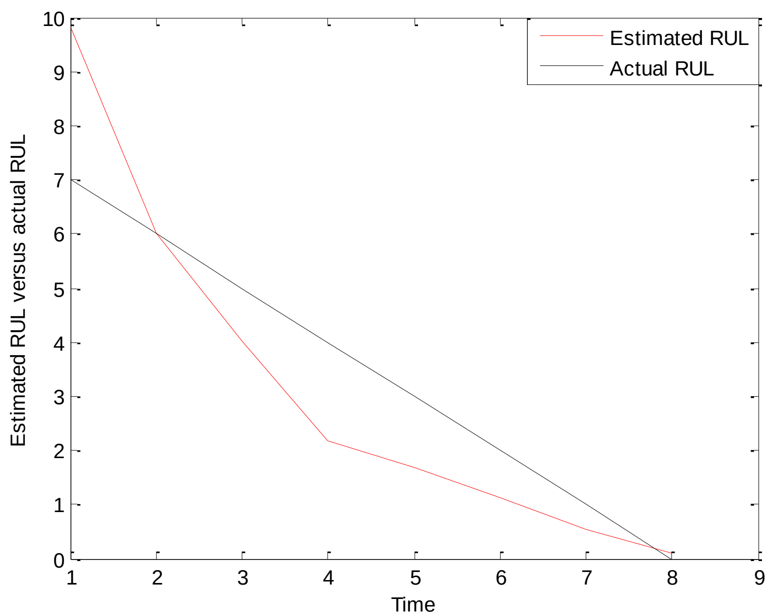


Figure 3-3: Actual RUL versus estimated RUL for equipment 11

The previous steps of the methodology constitute three main phases; training, updating, and RUL calculation. Steps 2, 3, and 4 constitute the training phase while steps 5 is the updating phase, and step 6 is the RUL calculation phase. The pseudocodes of these three phases are explained in two algorithms. Algorithm 1 is depicted in Figure 3-4, it illustrates the training phase. Algorithm 2 is depicted in Figure 3-5, it illustrates the updating and the estimation of the RUL since it is calculated instantaneously based on the updated survival curve.

**Algorithm 1: Training**

**Input:** $\boldsymbol{\Omega^{SL} = O\left(i, t_{F_i} \leq t_S, Z_{i,t_{F_i} \leq t_S}\right)}$, $\boldsymbol{\Omega^{LL} = O\left(i, t_{F_i} > t_S, Z_{i,t_{F_i} > t_S}\right)}$, and $\boldsymbol{i = \{1, \dots, T\}}$

**Output:** Set of generated patterns $\boldsymbol{P}$, baseline survival curve $\boldsymbol{S_b(t)}$, pattern survival curves $\boldsymbol{S_{p_j}(t)}$, and $\boldsymbol{j \in \{1, \dots . |P|\}}$

$\boldsymbol{S_b(t)}$ ← Estimation of baseline survival function using eq. (3.1), by considering $\boldsymbol{t_{F_i}}$.

$\boldsymbol{P = \left\{p_j, j \in \{1, \dots . |P|\}\right\}}$ ← Pattern generation using LAD classifier considering $\boldsymbol{Z_{i,t_{F_i} \leq t_S}}$ **and** $\boldsymbol{Z_{i,t_{F_i} > t_S}}$

**for** $\boldsymbol{j = 1}$ **to** $\boldsymbol{|P|}$

    **for** $\boldsymbol{i = 1}$ **to** $\boldsymbol{T}$

        $\boldsymbol{cov(p_j)}$← Using LAD classifier to obtain the coverage of pattern $\boldsymbol{p_j}$ considering $\boldsymbol{Z_{i,t_{F_i}}}$

    **return** $\boldsymbol{cov(p_j)}$

    $\boldsymbol{S_{p_j}(t)}$ ← Estimation of the survival function for the pattern $\boldsymbol{p_j}$ using eq. (3.2)

**return** $\boldsymbol{P, S_b(t), \left\{S_{p_j}(t) : j \in \{1, \dots . |P|\}\right\}}$

Figure 3-4: The pseudocode of the training algorithm

**Algorithm 2: Updating the survival curve and RUL calculation for the equipment $\boldsymbol{u}$ in the updating dataset**

**Input:** $\boldsymbol{\{O(u, t_k, Z_{u,t_k}), t_k = 1, 2, \dots t_{Fu})\}}$, $\boldsymbol{S_b(t), \left\{S_{p_j}(t) : j \in \{1, \dots . |P|\}\right\}, u = \{1, \dots, U\}}$

**Output:** Updated survival curve $\boldsymbol{S_{O(u,t_k,Z_{u,t_k})}(t)}$, and remaining useful life $\boldsymbol{RUL_u(t_k)}$

**for** $\boldsymbol{t_k = 1}$ **to** $\boldsymbol{t_{Fu}}$

    $\boldsymbol{j \in \{1, \dots . |P|\}}$ ← Using LAD diagnoser to obtain the set of covering pattern considering $\boldsymbol{Z_{u,t_k}}$

    $\boldsymbol{S_{O(u,t_k,Z_{u,t_k})}(t)}$← Updating the survival curve for the equipment $\boldsymbol{u}$ using eq. (3.3) or eq. (3.4).

    $\boldsymbol{RUL_u(t_k)}$ ← RUL estimation for the equipment $\boldsymbol{u}$ using eq. (3.7).

**return** $\boldsymbol{S_{O(u,t_k,Z_{u,t_k})}(t), RUL_u(t_k)}$

Figure 3-5: The pseudocode of the updating and RUL estimation algorithm

The next section discusses a comparison between the *MRUL* values calculated by using the proposed methodology and those obtained when using PHM model. A case study is presented.

## 3.5 Case Study

### 3.5.1 Turbofan engine dataset

To validate the proposed methodology, we use the turbofan engine dataset (C-MAPSS dataset) that is available on the website of NASA prognostic data repository (Saxena, et al. 2008). The training dataset is collected from a set of 260 aircraft engines. In this dataset, each row represents an observation taken during a single operational cycle. Each observation is represented by the corresponding equipment identity, the age of the equipment in cycles, in addition to twenty-four covariates. The first, the second, and the third covariates are three operational settings while all the remaining covariates represent sensor measurements.

**Training dataset.** We have 260 equipment (engines) $(i = 1,2, \ldots 260)$, each equipment fails after a number of cycles. The equipment that fail before the 206th cycle (which is the *MTTF*) is considered as *SL* while the ones that fail after that time are considered as *LL* equipment. The training data is splitted into two classes. The first class contains the failure observations collected from equipment that fail before the mean time to failure (206 cycles) while the second one contains the failure observations collected from the equipment that fail after this time as listed in Table 3.9. Accordingly, we have 154 *SL* equipment and 106 *LL* equipment.

Table 3.9: Representation of the training observations for the 260 equipment

| | | |
|---|---|---|
| Training dataset $\Omega$ | *SL* observations $\Omega^{SL}$ (154 Equipment) | $O(i, t_{Fi} \leq 206, Z_{i,t_{Fi} \leq 206})$ |
| | *LL* observations $\Omega^{LL}$ (106 Equipment) | $O(i, t_{Fi} > 206, Z_{i,t_{Fi} > 206})$ |

### 3.5.2 Pattern generation

The *SL* and *LL* patterns are generated from the twenty-four covariates of the training observation using two-class LAD. There are sixty-eight generated patterns (fourty-one *SL* and twenty-seven *LL* patterns). A sample of the generated patterns, and the corresponding covered equipment, are listed in Table 3.10.

Table 3.10: The generated *SL* and *LL* patterns

| Generated *SL* Patterns | Identities of the covered equipment |
|---|---|
| $p_4$ | 8, 9, 26, 28, 29, 36, 63, 67, 70, 74, 89, 111, 130, 178, 182, 194, 197, 208, 212, 234, 247, 250, 253 |
| $p_{17}$ | 8, 16, 26, 45, 55, 63, 70, 72, 129, 130, 137, 138, 166, 187, 198, 208, 210, 212, 229, 247, 250, 259 |
| $p_{21}$ | 1, 6, 8, 14, 16, 26, 38, 55, 58, 63, 70, 120, 130, 133, 136, 137, 138, 154, 166, 187, 208, 212, 247, 250 |
| $p_{23}$ | 8, 26, 44, 63, 67, 70, 71, 77, 87, 99, 122, 130, 132, 141, 152, 208, 212, 231, 247, 256 |
| $p_{30}$ | 8, 25, 26, 28, 34, 36, 44, 55, 63, 67, 70, 83, 89, 114, 117, 130, 162, 176, 179, 182, 188, 194, 197, 208, 212, 217, 225, 241, 247, 253, 258 |
| $p_{39}$ | 25, 28, 34, 36, 60, 83, 89, 92, 98, 107, 117, 165, 176, 179, 182, 194, 197, 205, 213, 217, 225, 238, 253, 258 |
| Generated *LL* Patterns | Identities of the covered equipment |
| $p_{43}$ | 4, 15, 31, 32, 75, 80, 100, 112, 113, 156, 161, 164, 168, 181, 183, 216, 218, 220, 240, 248 |
| $p_{53}$ | 13, 23, 43, 47, 48, 61, 73, 81, 84, 85, 94, 110, 123, 124, 135, 150, 159, 171, 180, 206, 215, 223, 228, 243, 245, 257, 260 |
| $p_{55}$ | 23, 43, 81, 84, 88, 91, 94, 109, 110, 123, 135, 144, 147, 150, 159, 171, 180, 183, 204, 206, 215, 228, 257 |
| $p_{58}$ | 4, 12, 15, 31, 32, 37, 64, 75, 112, 116, 121, 128, 146, 153, 155, 158, 181, 216, 218, 239 |
| $p_{60}$ | 23, 41, 47, 61, 81, 94, 104, 105, 106, 110, 124, 131, 134, 145, 148, 159, 160, 171, 180, 206, 223, 242, 243, 245 |
| $p_{70}$ | 23, 43, 81, 84, 88, 94, 109, 110, 123, 135, 144, 150, 159, 171, 180, 183, 204, 206, 215, 228, 254, 257 |

The survival curves for the generated *SL* and *LL* patterns are shown in Figure 3-6. As expected, it is noticed from the figure that the survival curves of the *SL* (*LL*) patterns are below (above) that of the baseline.

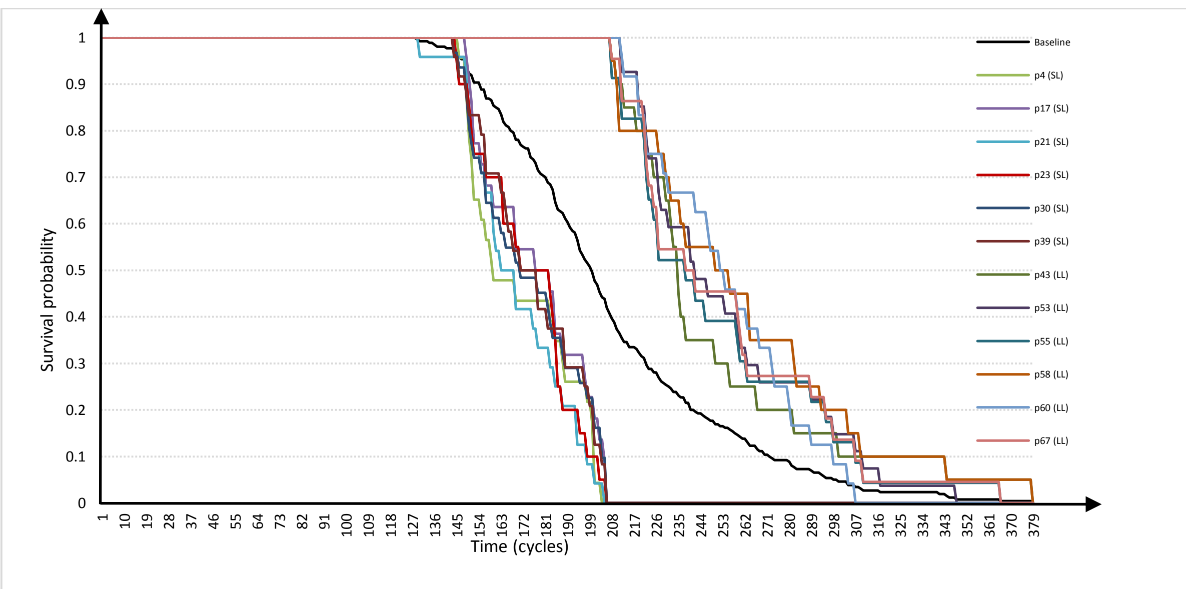


Figure 3-6: Survival curves for some of the generated *SL* and *LL* patterns

***Updating dataset.*** We employ the data collected from another set of similar equipment (259 equipment) from C-MAPSS dataset to update the survival curve and to calculate the *MRUL*. The proposed LAD prognostic methodology is compared to Cox proportional hazard model (Cox PHM) as one of the common prognostic techniques. Cox PHM has been employed in the field of CBM prognostic due to its capability of reflecting the effect of covariates on the equipment age. The survival function in Cox PHM is expressed as (Hosmer Jr, et al. 2011):

$$S(t, Z) = S_0(t)^{\exp(\beta Z)} \tag{3.8}$$

where $Z$ is a vector of covariates, $S_0(t)$ is the baseline survival function which characterizes the effect of the age of the equipment on its failure time, and $exp(\beta Z)$ is a function of weighted covariates that affect the failure time.

In what follows, we compare the results of *MRUL* estimation using the proposed LAD methodology and the PHM prediction model. Since each equipment in the updating dataset has different number of operational cycles, it is important to have a good performance measure to validate the prognostic methodology, we use the *root mean squared error* (*RMSE*) as a significant error measurement criterion. The *RMSE* for the *MRUL* estimation of the $u^{th}$ equipment is calculated as follows:

$$RMSE(u) = \sqrt{\frac{\sum_{t_k=1}^{t_{Fu}} [actual\ RUL\ \ (t_k) - estimated\ MRUL(t_k)]^2}{N_{Fu}}} \tag{3.9}$$

where $N_{Fu}$ is the actual number of operational cycles until the failure of the $u^{th}$ equipment, and $u = 1, \ldots, 259$.

As an example, the *RMSE* values of estimated *MRUL* for some of the equipment using *LAD-Model 1*, *LAD-Model 2*, and PHM prediction model, are listed in Table 3.11. The underlined bold results indicate that the corresponding model gives better results than the others (better result means smallest *RMSE* values).

Table 3.11: The results for LAD prognostic models and PHM prediction model

| Equipment Identity | Time To Failure (TTF) in cycles | RMSE using *LAD Model 1* | RMSE using *LAD Model 2* | RMSE using PHM |
|---|---|---|---|---|
| 4 | 196 | 29.35168 | **<u>27.95178</u>** | 54.74630 |
| 40 | 220 | 34.36686 | **<u>31.60582</u>** | 38.80056 |
| 56 | 225 | **<u>26.18971</u>** | 28.52649 | 30.14525 |
| 65 | 378 | 58.95770 | **<u>54.54994</u>** | 69.03294 |
| 67 | 232 | 28.66452 | 26.75004 | **<u>24.66802</u>** |
| 73 | 229 | 25.02456 | **<u>23.56540</u>** | 29.17804 |
| 121 | 126 | 15.39275 | **<u>14.66081</u>** | 19.34902 |
| 128 | 218 | 30.87016 | **<u>28.83593</u>** | 36.57858 |
| 150 | 230 | **<u>29.38356</u>** | 32.02670 | 30.43499 |
| 176 | 223 | 28.03367 | **<u>25.79815</u>** | 35.84996 |
| 189 | 221 | 24.89997 | **<u>23.79137</u>** | 36.05095 |
| 207 | 226 | 27.22564 | **<u>25.21357</u>** | 30.25787 |
| 219 | 234 | 28.59781 | 27.03575 | **<u>25.18294</u>** |
| 234 | 236 | **<u>20.97568</u>** | 21.32425 | 29.17945 |
| 258 | 229 | **<u>21.58249</u>** | 22.45859 | 27.83652 |

### 3.5.3 Result validation using two-phase Friedman test

Based on the results of *MRUL* estimation for the 259 equipment, we need to properly determine whether the performance of the proposed LAD prognostic models outperform that of PHM prediction model or not. In this paper, the question that need to be answered is; is any model ranked consistently higher or lower than the others? In other words, does any of the models perform consistently better or worse RUL estimation?

Since each equipment has different number of operational cycles, we need a suitable non-parametric statistical test to report the comparison of all models by considering that difference. Friedman test is one of the recommended tests in such situations (Daniel 1990). Friedman test can be employed even if the underlying distribution of the variable to be tested (in this case it is the *RMSE*) is asymmetrical and unknown. In two-phase Friedman test, we first test the hypothesis that all the prognostic models applied are statistically equivalent or not, afterwards post tests must be carried out to report the best model.

**Phase 1: Test hypothesis**

The *RMSE* is calculated for each equipment using the three prognostic models; *LAD-Model 1*, *LAD-Model 2*, and *PHM*. In this experiment, the results obtained from 259 equipment rate those three models. Friedman test is based on ranks rather than on the original raw data. It starts by assigning ordered ranks to the *RMSE* values calculated for each equipment, from the smallest to the largest.

The null hypothesis in the Friedman test states that all the prognostic models have the same mean *RMSE*. It is formulated as:

$$H_0: \mu_{LAD\_Model\ 1} = \mu_{LAD\_Model\ 2} = \mu_{PHM}$$

And the alternative hypothesis is formulated as:

$$H_a: Not\ all\ RMSE\ means\ \{\mu_{LAD\_Model\ 1}, \mu_{LAD\_Model\ 2}, \mu_{PHM}\}are\ equal$$

The test statistic is given by the following equation (Daniel 1990):

$$F_r = \frac{12}{UK(K+1)} \sum_{j=1}^{K} R_j^2 - 3U(K+1) \tag{3.10}$$

Where $R_j$ is the total rank sum for the prognostic model $j$ (*j=1,..K*), *K* is the total number of the prognostic models (*K*=3 in this case), and *U* is the total number of equipment in the updating dataset (*U=259*). In this work, the values of the total ranks for the three models are listed in the second column of Table 3.12. The calculated mean ranks for the three models are listed in the third column of Table 3.12. The calculated significance level ( *p-value*) is given by:

$$p - value = Pr(\chi^2_{K-1} \geq F_r) = Pr(\chi^2_{3-1} \geq 60.33) = 0$$

And the declared test significant level is calculated as: $\chi^2_{K-1,\alpha} = \chi^2_{2,0.05} = 5.9914$. It is noticed that the *p-value* is too small, the test is said to be significant, and accordingly the null hypothesis is rejected. Rejecting $H_0$ can be interpreted to mean that some prognostic models tend to have larger or smaller *RMSE* means than the other. Consequently, an appropriate post-hoc multiple comparisons test (pairwise Friedman test) is performed.

**Phase 2: Pairwise comparisons**

After the rejection of the null hypothesis, we are confident that one of the three prognostic models including PHM prediction model outperforms the other two ones. After performing the previous procedures, there is no indication as to which models have better performance and which ones have a similar performance. We want to know whether the performance of LAD prognostic models exceed that of PHM model or not.

A pairwise Friedman test is performed to compare the *RMSE* estimated using PHM and those estimated using LAD prognostic models (one comparison between PHM model and each of LAD prognostic models). One hypothesis test for each comparison is formulated and a decision should be made regarding the rejection or acceptance of that hypothesis. In the comparison of LAD prognostic *Model j* and *PHM,* the null and the alternatives hypotheses are formulated as:

$$H_0: \mu_{PHM} = \mu_{LAD\_Model\ j}$$

$$H_a: \mu_{PHM} > \mu_{LAD\_Model\ j}$$

First, we compute the absolute difference between the rank sums $\left|R_{LAD\_Model\ j} - R_{PHM}\right|$ and if this difference exceeds the value $d_{\alpha_F}\sqrt{UK(K+1)/6}$, the null hypothesis is rejected. Where $d_{\alpha_F}$ is the $100(1-\alpha_F)^{th}$ of the standard normal distribution (Gwet 2011). The results and the final decision discussed verbally are presented briefly in Table 3.12.

It is noticed from Table 3.12 that the results obtained by using *LAD Model 2* outperform those obtained when using *LAD Model 1* and *PHM Model*. This can be noticed through the underlined bold value, in the fourth column of Table 3.12. Each value in this column is the difference between the total rank of *LAD Model j* and the total rank of *PHM*.

Table 3.12: Friedman test to validate the proposed LAD methodology

| Prognostic Model | Total rank $R_j$ | Mean rank | $\lvert R_j - R_{PHM} \rvert$ | Is $\lvert R_{LAD\_Modelj} - R_{PHM} \rvert$ exceeding $d_{\alpha_F}\sqrt{UK(K+1)/6}$? | Acceptance of the null hypothesis? | Significant difference? |
|---|---|---|---|---|---|---|
| *LAD Model 1* | 539 | 2.08 | 55 | Yes | No | **<u>Yes</u>** |
| *LAD Model 2* | 421 | 1.63 | **<u>173</u>** | Yes | No | **<u>Yes</u>** |
| *PHM Model* | 594 | 2.29 | | | | |
| Test Statistic $F_r$ | 60.33 | | | | | |
| $\chi^2_{k-1,\alpha}$ | 5.9914 | | | | | |
| P-value | 0 | | | | | |
| $d_{\alpha_F}$ | 1.96 | | | | | |
| $d_{\alpha_F}\sqrt{UK(K+1)/6}$ | 44.61 | | | | | |

Therefore, our final decision after the comparisons listed in Table 4.12 can be interpreted as follow:

- *There is a significant difference between the calculated RMSE values when using LAD Model 1, and that of using PHM Model.*
- *There is a significant difference between the calculated RMSE values when using LAD Model 2, and that of using PHM Model. Morever, LAD Model 2 is competitive to both of LAD Model 1 and the PHM Model.*

## 3.6 Conclusion

In this paper, we developed a two-class LAD prognostic methodology. The proposed methodology updates the survival (reliability) curve of an observed equipment and estimate its RUL. The survival curve of each observation collected from the equipment is updated by using a set of generated patterns covering that observation.

The application of the two-class LAD in fault prognosis in CBM is done in a number of steps. In the first step, the historical data that consist of training and updating datasets is prepared. The training dataset is collected from a set of equipment, and contains the failure times and the corresponding covariates. The equipment in the training dataset are classified into two class (*short life* and *long life*). This is carried out by using the *MTTF* to separate the time to failures into two categories. The updating dataset is collected from another set of similar equipment, and also contains the aging times and the corresponding covariates. In the second step, the baseline survival function is estimated from the failure times of all equipment using Kaplan-Meier (KM) estimator. In the third step, the two-class LAD classifier is used to generate patterns that differentiate between the short life and long life equipment, in the training dataset. In the fourth step, the updating observations collected from each equipment is used to update its survival curve, based on the patterns that are covering the most recent observation. One of two prognostic formulas is used to update that survival curve. In the final step of the methodology, the RUL of the equipment is calculated based on its updated survival curve.

It is clearly shown from the obtained results that the proposed LAD prognostic methodology outperforms PHM prediction model in particular when applying *LAD Model 2* (there is a higher significant difference in the performance when using *LAD Model 2*). This is may be attributed to the fact that the survival curve in *LAD Model 1* is updated by assigning a weight for the baseline curve equal to the weight of the survival curve of each pattern covering the updating observation, although the coverage of each pattern is less than the coverage of the baseline. In *LAD Model 2*, the survival curve is updated by assigning a weight for the baseline curve greater than the weight of the survival curve of each pattern covering the observation.

As a final conclusion, the proposed prognostic methodology is promising estimation technique for the estimation of RUL in the field of CBM prognostics since it is not based on any thresholding strategy to estimate the RUL. Unlike many statistical prognostic methods, LAD prognostic methodology does not need to satisfy any statistical assumptions, and can perform well when the covariates are highly correlated as the case of many practical situations.

One of our future research trends is to study the effect of taking all the equipment's history in order to generate the patterns. We are also studying the effect of generating non pure patterns on the performance of LAD prognostic methodology. Another trend is to consider a weight for each pattern in the updating formula that reflect the true coverage (the number of the covered observations in the training dataset). One of our research challenge is to extend the proposed methodology to the case of multiple failure modes (multi-fault prognosis).

## 3.7 References

Alexe, S., Blackstone, E., Hammer, P. L., Ishwaran, H., Lauer, M. S., & Pothier Snader, C. E. (2003). Coronary risk prediction by logical analysis of data. *Annals of Operations Research*, *119*, 15–42.

Alexe, G., Alexe, S., Bonates, T. O., & Kogan, A. (2007). Logical analysis of data—the vision of Peter L. Hammer. *Annals of Mathematics and Artificial Intelligence*, *49*, 265–312.

Banjevic, D., & Jardine, A. (2007). Remaining useful life in condition based maintenance: Is it useful? In *Modelling in industrial maintenance and reliability* (p. 7).

Bennane, A., & Yacout, S. (2012). LAD-CBM; new data processing tool for diagnosis and prognosis in condition-based maintenance. *Journal of Intelligent Manufacturing*, *23*, 265–275.

Bores, E., Hammer, P. L., Ibaraki, T., Kogan, A., Mayoraz, E., & Muchnik, I. (2000). An implementation of logical analysis of data. *IEEE Transactions on Knowledge and Data Engineering*, *12*, 292–306.

Caesarendra, W., Widodo, A., & Yang, B. S. (2010). Application of relevance vector machine and logistic regression for machine degradation assessment. *Mechanical Systems and Signal Processing*, *24*, 1161–1171.

Crama, Y., Hammer, P. L., & Ibaraki, T. (1988). Cause-effect relationships and partially defined Boolean functions. *Annals of Operations Research*, *16*, 299–325.

Daniel, W. W. (1990). *Applied nonparametric statistics* (2nd Ed.). Boston: PWS-KENT Pub.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification. A Wiley-Interscience publication*. New York: Wiley.

Elsayed, E. A. (2012). *Reliability engineering*. London: Wiley.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer Series in Statistics.

Guo, C., & Ryoo, H. S. (2012). Compact MILP models for optimal and Pareto-optimal LAD patterns. *Discrete Applied Mathematics*, *160*, 2339–2348.

Gwet, K. L. (2011). *The practical guide to statistics: Applications with excel, R, and calc*. Gaithersburg, MD: Advanced Analytics, LLC.

Hamada, M. (2005).Using degradation data to assess reliability. *Quality Engineering*, *17,* 615–620.

Hammer, P. L., Kogan, A., Simeone, B., & Szedmák, S. (2004). Pareto optimal patterns in logical analysis of data. *Discrete Applied Mathematics*, *144*, 79–102.

Hammer, P. L., & Bonates, T. O. (2006). Logical analysis of data—an overview: From combinatorial optimization to medical applications. *Annals of Operations Research*, *148*, 203–225.

Heng, A., Tan, A. C. C., Mathew, J., Montgomery, N., Banjevic, D., & Jardine, A. K. S. (2009). Intelligent condition-based prediction of machinery reliability. *Mechanical Systems and Signal Processing*, *23*, 1600–1614.

Hosmer, D. W, Jr, & Lemeshow, S. (2011). *Applied survival analysis: Regression modeling of time to event data* (Vol. 618). New York: Wiley.

Jardine, A., Joseph, T., & Banjevic, D. (1999). Optimizing condition based maintenance decisions for equipment subject to vibration monitoring. *Journal of Quality in Maintenance Engineering*, *5*, 192–202.

Jardine, A. K. S., Lin, D., &Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, *20*, 1483–1510.

Kim, H.-E., Tan, A. C., Mathew, J., & Choi, B.-K. (2012). Bearing fault prognosis based on health state probability estimation. *Expert Systems with Applications*, *39*, 5200–5213.

Klein, J., & Moeschberger, M. (1997). *Survival analysis: Techniques for censored and truncated data*. New York: Spring.

Kothamasu, R., Huang, S. H., & VerDuin, W. H. (2006). System health monitoring and prognostics—a review of current paradigms and practices. *The International Journal of Advanced Manufacturing Technology*, *28*, 1012–1024.

Kronek, L. P., & Reddy, A. (2008). Logical analysis of survival data: Prognostic survival models by detecting high-degree interactions in right-censored data. *Bioinformatics*, *24*, i248–i253.

Le Son, K., Fouladirad, M., Barros, A., Levrat, E., & Iung, B. (2013). Remaining useful life estimation based on stochastic deterioration models: A comparative study. *Reliability Engineering and System Safety*, *112*, 165–175.

Liao, H., Zhao, W., & Guo, H. (2006). Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model (pp. 127–132).

Mortada, M.-A., Yacout, S., & Lakis, A. (2013). Fault diagnosis in power transformers using multi-class logical analysis of data. *Journal of Intelligent Manufacturing*. Doi: 10.1007/s10845-013-0750-1.

Mortada, M.-A., Yacout, S., & Lakis, A. (2011). Diagnosis of rotor bearings using logical analysis of data. *Journal of Quality in Maintenance Engineering*, *17*, 371–397.

Pintilie, M. (2006). *Competing risks: A practical perspective* (Vol. 58). New York: Wiley.

Ryoo, H. S., & Jang, I. Y. (2009). Milp approach to pattern generation in logical analysis of data. *Discrete Applied Mathematics*, *157*, 749–761.

Saha, B., & Goebel, K. (2008).Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques (pp. 1–8).

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation (pp. 1–9).

Schwabacher, M., & Goebel, K. (2007). A survey of artificial intelligence for prognostics. *Paper presented at the Artificial Intelligence for Prognostics-AAAI Fall Symposium, November 9–11* (pp. 107–114). Arlington, VA.

Tan, A., Heng, A. S.Y., & Mathew, J. (2009).Condition-based prognosis of machine health. In *Proceedings of the 13th Asia-Pacific Vibration Conference* (pp. 1–10). University of Canterbury.

Tian, Z., Lin, D., & Wu, B. (2012). Condition based maintenance optimization considering multiple objectives. *Journal of Intelligent Manufacturing*, *23*, 333–340.

Tian, Z., Wong, L., & Safaei, N. (2010). A neural network approach for remaining useful life prediction utilizing both failure and suspension histories. *Mechanical Systems and Signal Processing*, *24*, 1542–1555.

Vachtsevanos, G. J., Lewis, F. L., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent fault diagnosis and prognosis for engineering systems*. London: Wiley.

Wang, W. (2007). A prognosis model for wear prediction based on oilbased monitoring. *Journal of the Operational Research Society*, *58*, 887–893.

Widodo, A., & Yang, B. S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, *21*, 2560–2574.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Los Altos, CA: Morgan Kaufmann.

# CHAPTER 4 ARTICLE 2 : PATTERN-BASED PROGNOSTIC METHODOLOGY FOR CONDITION BASED MAINTENANCE USING SELECTED AND WEIGHTED SURVIVAL CURVES

Ahmed Ragab, Soumaya Yacout, Mohamed-Salah Ouali, and Hany Osman

Submitted to: IEEE TRANSACTIONS ON RELIABILITY

Manuscript Number: TR-2014-464

## 4.1 Abstract

This paper proposes a pattern-based prognostic methodology which combines Logical Analysis of Data (LAD) as an event-driven diagnostic technique, and Kaplan-Meier (KM) estimator as a time-driven technique. LAD captures the effect of the instantaneous conditions on the health state of the monitored system, while KM estimates the baseline survival curve that reflects the effect of the aging, based on the observed historical failure times. LAD is used to generate a set of patterns from the observed values of the covariates that represent the operating conditions and the condition indicators. A pattern selection procedure is carried out to select the set of significant patterns form all the generated patterns. A survival curve is estimated, for each subset of observations covered by each selected pattern, by using KM estimator. A weight that reflects the coverage of each patterns, is assigned to its survival curve. Given a recently collected observation, the survival curve of the monitored system is updated based on the patterns that cover that observation. The updated curve is used to predict the remaining life of the system. The methodology is validated using a common dataset in prognostics; the turbofan degradation data that is available at NASA prognostic data repository. Moreover, the methodology is compared to two machine learning regression techniques.

## 4.2 Introduction

Condition based maintenance (CBM) aims at avoiding unnecessary preventive maintenance tasks by taking preventive replacements only when there are signs of potential failure of the monitored system [1]. Thus, to make the appropriate decision in CBM, a fault prognostic strategy should be deployed [2]. Researches supported by international agencies, industry and academia, focus on developing intelligent CBM prognostic systems [3, 4]. Basically, maintenance decisions in CBM prognostics are based on up-to-date collected data (i.e. observations) from the monitored system [5]. These data consist of condition monitoring data and event data [6]. Condition monitoring data allow determining the system's health condition. Event data provide the information on what happened to the system (failure, installation, overhaul, repair, inspection, etc.) [7]. When event data represent the failure mechanisms of the system, they are called lifetime data. They are used to assess the reliability (also called survival) function of the system by using either parametric or non-parametric estimation techniques [8]. The survival function allows to predict the remaining useful life (RUL) of the system, defined as the time left for a system in operation before the occurrence of complete failure [9]. Reliability-based prognosis methods provide some modeling options to process the collected data and predict the RUL of industrial systems [3].

From practical point of view, estimating the RUL of a monitored system, by using the condition monitoring data, is a particular research area since there is no clear and known relationship between the condition measurements and the prediction of RUL. In fact, the estimated survival function does not reflect explicitly the effects of instantaneous conditions of each single system on its health state. Thus, the estimated RUL is not related to those system's specific conditions, and as such are not very accurate [10]. In many industrial situations, the prognostic of RUL still relies on reliability experts who may have significant experience about the failure and the degradation states of the system. However, they do not have systematic methodologies that can predict the RUL using instantaneous conditions of the monitored system, and the experience is becoming hard to accumulate due to the complexity of the monitored system [3].

The accuracy of current prognostic methods is limited by some common challenges, as follows [3, 6, 9, 10]:

- The operating conditions and the condition indicators may be time-varying or intercorrelated. The challenge is to deal with such variation and correlation without making any statistical assumption.
- The limited number of both event data and condition monitoring data. This can happen when the data are collected from similar but not identical systems, supplied from different manufacturers. The challenge is to design effective prognostic methods that can fully utilize all the data from systems with different specifications.
- Noisy and missing data. The challenge is to enlarge the capacity of a given prognostic method to deal with both noisy and missing data.
- Current reliability-based prognostic methods need predetermined threshold to predict the failure time of the monitored system. When such threshold is not known, the prediction accuracy will be affected accordingly. The challenge is to design a prognostic method without considering any threshold values.

To deal with some of the above challenges, a powerful pattern-based machine learning approach called Logical Analysis of Data (LAD) was applied in CBM [11]. From the perspective of the CBM decision makers, LAD is used as a supervised learning technique to automatically generate, from condition monitoring data, interpretable patterns and translate them into diagnostic rules, without any statistical treatments or assumptions, even if the data are highly correlated or time varying [12, 13]. Moreover, LAD deals with noisy or missing data, as reported in [14, 15]. From the reliability analysis perspective, Kaplan-Meier (KM) is commonly applied as a non-parametric technique to provide an estimate for the survival function, based on lifetime data collected from a set of systems [16]. However, KM reflects only the effect of age on the systems' health state. It does not reflect explicitly the effects of the instantaneous conditions on the health state of the system.

The aim of this paper is to address some of the above common challenges in CBM prognosis. It proposes a reliability-based prognostic methodology based on LAD and KM, in order to predict the RUL of a monitored system. The proposed methodology uses both of LAD as an event-driven diagnostic technique and KM reliability modeling as a time-driven estimation technique. LAD offers some important advantages, particularly when it deals with highly correlated covariates without any statistical treatment or preprocessing [17]. The advantage of using KM is that, it is a non-parametric estimation method which does not need any statistical assumptions about the distribution of lifetime data [16]. KM is used to estimate the baseline reliability curve. The proposed methodology is inspired from the approach called logical analysis of survival data (LASD) that was presented in the field of medical prognosis [18]. However, LASD has two major limitations when updating the survival function. The first one is that, the survival curve of each pattern has a weight that is equal to the weight of the baseline survival curve, although none of the generated patterns can cover the same number of observations that are covered by the baseline curve. The second limitation is that, the survival curve for the recent observation is updated by averaging the baseline curve and the survival curves of the patterns that cover that observation, while ignoring the updated survival curve that was obtained from the previous observation.

The above two limitations are already addressed in the prognostic methodology that is proposed in [13]. Furthermore, a single measure that calculates the RUL based on the updated survival curve is provided. However, that methodology considers only the condition monitoring data that are collected at/or immediately before the system failure. This limitation means that all the bulk of information embedded in previous observations are ignored. In order to address this specific limitation, three modifications are introduced in this paper. They are summarized as follows:

- All condition monitoring data collected from a group of systems during their life spans are considered (both normal and failure observations).
- Since the number of generated pattern may be very large, pattern selection procedure is used in order to select the set of significant patterns form all the generated patterns.
- Moreover, a weight that reflects the coverage of each pattern, thus its importance, is assigned to its survival curve.

The proposed enhanced methodology comprises training and updating phases. In the training phase, a subset of observations collected from a group of similar systems are used. Each observation consists of the time and the condition monitoring data in the form of the covariates that represent the operating conditions and the condition indicators. For each observation, the time can be operational time or failure time. Given the historical lifetime data of all systems, the baseline survival curve is estimated by KM technique. Historical condition monitoring data are fed into two-class LAD classifier to generate a set of patterns that reflects the effect of the covariates on the systems' health state. The generated patterns represent the interactions among the operating conditions and the condition indicators. A pattern selection procedure is carried out to remove the redundant patterns and to select the significant ones. For each selected pattern, a survival curve is obtained by using KM estimation technique, by considering only the observations that are covered by that pattern. At the end of the training phase, the knowledge is discovered in the form of the baseline survival curve and a set of survival curves for the selected patterns (one survival curve for each selected pattern).

In the updating phase, the knowledge discovered in the training phase is used to estimate, at a certain time, the health state of a given system, different from those used in the training phase. The latest observation collected from this system allows updating its survival curve. This can be done by using the patterns that are covering the recently collected observation. The covering patterns reflect the effect of the instantaneous conditions on the system's health state. After collecting the subsequent observation, the former updated curve for that system is updated again based on the new covering patterns for that observation, and so on. Thus, each system has its own updated survival curve that reflects its specific working condition. Based on the new collected observation and the corresponding updated survival curve, the RUL of the system is estimated.

This paper is divided into five sections. Section 4.3 presents LAD approach along with the definition and characteristics of the pattern, and the pattern generation and selection techniques. Section 4.4 explains the enhanced prognostic methodology and its procedure in details. Section 4.5 shows the obtained results by using a common case study in the field of prognostics. It presents a comparison between the proposed methodology and two common machine learning regression techniques: artificial neural networks (ANNs) and support vector regression (SVR). Section 4.6 concludes the paper.

## 4.3 Logical Analysis Of Data

### 4.3.1 The approach

Logical Analysis of Data is a supervised pattern generation and classification technique that was introduced in [19]. Originally, LAD was used as two-class classification method, that is a *dichotomizer* which is used to extract knowledge from the dataset that contain two classes of observations [14]. Each observation is represented as a vector of binary or numerical values of certain characteristic features or covariates. LAD has been evolved as an effective diagnostic technique which relies on extracting patterns from the training data, by using combinatorial and optimization methods [20]. The extracted patterns characterize and differentiate between the two classes of positive and negative observations, in the training dataset [17]. Each pattern represents the interactions between covariates for either the set of positive or the set of negative observations [21]. Accordingly, LAD can be used as pattern-based classifier for the new observations that are not included in the training dataset. It was applied on a various classification problems and the results in [14, 17] indicate that LAD's classification accuracy is comparable and often superior to other classification methods.

Two-class LAD is based on Boolean theory, it is composed of three stages; data binarization, pattern generation, and theory formation [14]. The binarization stage involves the transformation of numerical data to binary data using a binarization technique that transforms the numerical covariates' values $z_1, z_2, \ldots, z_r$ into a set of binary attributes $a_1, a_2, \ldots, a_m$, where $m > r$. A Boolean function $f(a_1, a_2, \ldots, a_m)$ is a mapping $[0,1]^m \rightarrow [0,1]$. A partially defined Boolean function (PDBF) is given by a set of $m$ dimensional 0-1 vectors and is denoted by $(\Omega^+, \Omega^-)$, where $\Omega^+ \subseteq [0,1]^m$ is the set of *positive* binary observations, and $\Omega^- \subseteq [0,1]^m$ is the set of *negative* binary observations. The function $f(A)$ is called an *extension* of the PDBF $(\Omega^+, \Omega^-)$, it is defined as:

$$f(A) = \begin{cases} 1 & \text{if} \quad A \in \Omega^+ \\ 0 & \text{if} \quad A \in \Omega^- \end{cases}$$

where $A = (a_1, a_2, \dots, a_m)$ is the binary observation vector obtained after the binarization of the numerical observation vector $Z = (z_1, z_2, \dots, z_r)$. For more details about LAD approach, the binarization, and the PDBF, the interested readers are referred to [14, 17, 19, 21, 22].

### 4.3.2 Definition and properties of patterns

The patterns generation stage is the key building block of LAD approach. This stage is essential in identifying the positive and negative patterns from the training dataset of positive and negative binary observations. A *literal* is either a binary attribute $a_l$ or its negation $\bar{a}_l$, where $(l = 1,2, \dots m)$. A *term* $C$ is a conjunction of distinct literals, on condition that it does not contain both a literal and its negation. The term $C$ covers the binary vector $A$ if $C(A) = 1$. The term $C$ is represented geometrically as the Boolean subcube of the $m$ dimensional cube $[0,1]^m$, not necessarily included in $\Omega^+ \cup \Omega^-$, corresponding to the set of points that are covered by $C$, denoted by $S(C)$, while $S(C) \cap \Omega^+ \cup \Omega^-$ is called coverage of $C$, denoted by $Cov(C)$ [23].

A *term* $C$ is called positive (or negative) *pattern* of the PDBF $(\Omega^+, \Omega^-)$ if: $C(A) = 0$ for every $A \in \Omega^-$ (or $A \in \Omega^+$) and $C(A) = 1$ for at least one vector $A \in \Omega^+$ (or $A \in \Omega^-$) [14]. Verbally, this means that the pattern can cover some observations from one class but cannot cover any observations from the opposite class. This is the concept of the strictly defined pattern. It is also called *pure* or *homogenous* pattern. Accordingly, a pure positive (negative) pattern is defined as an elementary conjunction of some of literals, that is true for at least one positive (negative) observation and false for all negative (positive) observations in the training dataset [19]. A pattern is said to cover a certain observation if it is true for that particular observation [14]. The set of observations covered by the pattern $p$ is denoted by $Cov(p)$.

The *degree* of a pattern indicates the number of literals involved in its definition. In other words, a pattern $p$ of degree $d$ is a conjunction of $d$ literals. A pattern $p_i$ is *strong* if there is no other pattern $p_k$ such that $Cov(p_i) \subset Cov(p_k)$. The accuracy of LAD classification depends on the type and characteristics of the generated patterns [20, 24]. The following subsection reviews some of the pattern generation approaches.

### 4.3.3 Pattern generation

There are three common approaches for pattern generation: enumeration-based [14], heuristic [22], and mixed integer and linear programming (MILP)-based approach [17, 20]. The MILP-based approaches are more accurate than the other approaches, give optimal solutions, and guarantee the generation of strong patterns, as reported in [20, 25].

In the MILP-based pattern generation method presented in [20], the objective is to maximize the number of positive (negative) observations that are covered by the generated positive (negative) patterns. The method guarantees that the generated patterns are optimal with respect to certain characteristics, for example the maximum coverage.

The generation procedure for each pattern, either positive or negative, is formulated as an MILP program. The procedure of positive pattern generation is identical to that of negative pattern generation. In what follows, the procedure for the generation of one pattern is stated briefly.

The decision variables of this formulation are defined first. Given a binarized training dataset composed of $m$ binary attributes, each generated pattern $p$ is associated with a Boolean vector $W = (w_1, w_2, \dots w_m, w_{m+1}, w_{m+2}, \dots w_{2m})$, whose size $2m$ is double that of the binary observation vector. For the generation of a positive (negative) pattern, the Boolean vector $Y$, whose number of elements equals the number of positive (negative) observations, is presented to indicate the coverage of the positive (negative) observations. The elements of the vectors $W$ and $Y$, and the degree $d$ of the generated pattern, are the decision variables in the MILP formulation.

A set of constraints is formulated on the values $w_l$ and $w_{m+l}$ to be mutually exclusive, i.e. $w_l + w_{m+l} \leq 1, \forall\, l = 1,2, \dots m$. An additional set of constraints is also formulated to ensure that the resulting pattern must be able to cover at least one observation in the corresponding class, and must not cover any observation in the opposite class.

After getting the optimal solution $(W^*,\ Y^*, d^*)$ to this MILP formulation, each binary attribute in the training dataset can be represented in the generated pattern as a literal or its negation. The elements of the vector $W^*$ are relative to the binary attributes such that if $w_l = 1$ then the literal $a_l$ is included in pattern $p$, and if $w_{m+l} = 1$ then its negation $\bar{a}_l$ is included in pattern $p$. Since the values $w_l$ and $w_{m+l}$ are mutually exclusive, the pattern $p$ cannot include both the literal $a_l$ and its negation $\bar{a}_l$ at the same time. The generated pattern $p$ is defined mathematically as a conjunction of the constituent literals as:

$$p \coloneqq \bigwedge_{\substack{w_l=1 \\ l\in\{1,..m\}}} a_l \bigwedge_{\substack{w_{m+l}=1 \\ l\in\{1,..m\}}} \bar{a}_l$$

The procedure for generating one pattern is repeated iteratively until each observation in the training dataset is covered by at least one generated pattern. The set of generated patterns is denoted by $P_{Gen}$. The interested readers can find more details about the MILP-based approaches in [17, 20, 25].

The existing pattern generation approaches tend to generate too many patterns, in order to cover all the observations in the training dataset. Some of the generated patterns may be redundant or cover a very small number of observations. Therefore we need a pattern selection approach in order to select the most significant patterns. This step improves the classification performance of LAD [26]. The following subsection discusses the pattern selection technique used in our prognostic methodology.

### 4.3.4 Pattern selection

The pattern selection procedure aims at improving the prediction accuracy of the two-class LAD decision model, by selecting the most significant patterns [26]. The selected patterns may result in a more stable performance in terms of being able to classify both positive and negative observations, due to their robustness to measurement errors. Consequently, this may result in more accurate RUL prediction in the proposed prognostic methodology, as we will explain in this paper. The pattern selection procedure finds the minimal subset of the patterns that cover all the observations, in order to discover as much knowledge from the training dataset. Each positive (negative) observation must be covered by at least one positive (negative) pattern. One way to eliminate the redundant patterns is to select them based on some criteria such as the coverage and the degree as in [27]. Alternatively, a subset of patterns can be selected from the generated patterns by solving a set covering problem (SCP) [14].

The pattern selection procedure used in our prognostic methodology is formulated as an SCP. The objective of the SCP formulation is to select the minimum number of patterns, while guaranteeing the coverage of all the training observations. In this SCP formulation, a Boolean vector

$X = (x_1, x_2, \dots x_{|P_{Gen}|})$ is defined such that the binary variable $x_g$ is equal to one (zero) if the generated pattern $p_g \in P_{Gen}$ is (not) selected in the required subset, where $g = 1, 2, \dots |P_{Gen}|$.

In order to select the minimum number of patterns with the maximum coverage, a covering constraint is defined for each observation in the training dataset in order to ensure the full coverage of all observations. Certainly, a set covering problem is solved for each class of observations included in the training dataset. The interested readers can find more details about the SCP formulation for pattern selection in [26].

The pattern generation and selection procedures used in this proposed methodology are carried out by using the software *cbmLAD* that was developed at *École Polytechnique de Montréal, Canada* [11]. The software uses the MILP-based approach presented in [17], for pattern generation. It also carries out the pattern selection procedure by formulating an SCP to find the minimal subset of the patterns that cover each observation in the training dataset.

## 4.4 The Proposed Reliability-Based Prognostic Methodology

### 4.4.1 Problem Statement

In this section, the proposed prognostic methodology is presented in details. Basically, the methodology combines KM estimation with LAD approach. KM is used to estimate the baseline survival curve, by using the historical lifetime data collected from a set of systems. However, the estimated baseline survival curve does not reflect the effect of the instantaneous conditions under which the systems are working. It reflects only the effect of the age on the health state of the systems. This limitation is overcome, by exploiting the condition monitoring data, in order to reflect explicitly the effect of the instantaneous conditions on the health state of the monitored system. Some systems fail before the mean time to failure (MTTF) due to working under severe operating conditions that lead to fast degraded performance, while others fail after the MTTF. We call the formers 'Short Life' systems, and the latter 'long Life' systems. The two-class LAD is used to generate patterns from the historical condition monitoring data that are collected from such set of monitored systems. The dataset consists of two categories of systems; 'Short Life' systems and 'long Life' systems. The generated patterns have the ability to discriminate between the 'short life' and the 'long life' systems. A survival curve is estimated for each generated pattern. The baseline survival curve, along with the survival curves of the patterns can be used to reflect the effect of the age and the instantaneous conditions on the health state of the monitored system.

### 4.4.2 Phases and steps of the methodology

The applicability of the two-class LAD in fault prognosis is carried out in seven steps. These steps constitute two phases; the first one is the training phase while the second one is the updating phase. The training phase contains the first five steps, while the updating phase consists of the last two steps. In what follows, the two phases and their steps are briefly explained, then the details of each step will be discussed. Figure 4-1 shows the schematic diagram of the proposed prognostic methodology.

**Training phase:** In step 1, the historical data that constitutes the training dataset is classified. This dataset consists of a set of observations collected from similar systems, and each observation contains the operational time and the corresponding covariates' values. The systems in the training dataset are classified into two classes; *short life* (*SL*) and *long life (LL*). This is done by separating the systems according to their corresponding failure times. If the system fails before the MTTF, it is called *SL*, otherwise it is called *LL* system. The updating dataset contains a set of observations collected from other similar systems, and contains the time and the corresponding covariates' values. In step 2, the baseline survival function is estimated from the failure times of all the systems in the training dataset using KM estimator. In step 3, the two-class LAD classifier is used to generate a set of *SL* and *LL* patterns that represent the hidden interactions between the covariates. Each pattern covers some observations of the systems in the corresponding class, and none of the observations in the opposite class. In step 4, a pattern selection procedure is then used in order to remove the redundant patterns. The most significant patterns are selected by solving a set covering problem that guarantees the coverage of all observations in the training dataset. In step 5, the survival curve for each of the selected *SL* and *LL* patterns, is estimated using KM estimator, by considering only the set of observations that are covered by each pattern, in the corresponding class.

**Updating phase:** In step 6 of the methodology, we consider a new monitored system that is similar to the ones used for training. The patterns that cover each collected observation of this system are identified by using LAD as a diagnostic technique. The KM survival curve is then updated according to the identified patterns, by calculating the weighted sum of the survival curves of these patterns and the KM survival curve. In step 7, the RUL of the monitored system is calculated, based on the updated survival curve, each time a new observation is collected. As a result of this updating procedure, the updated curve reflects the history of the operating and working condition of the monitored system, and consequently gives a better prediction of its RUL.

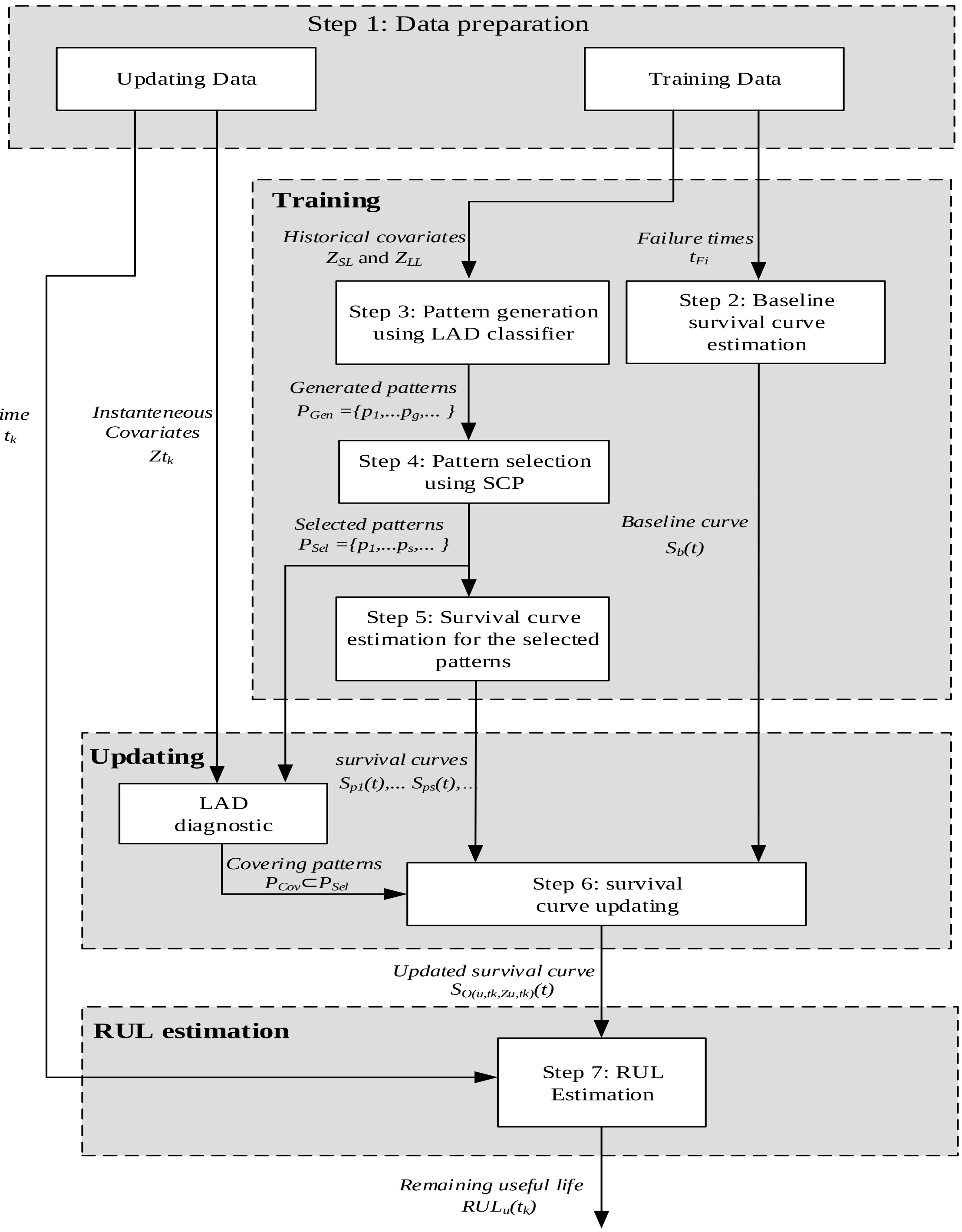


Figure 4-1: The diagram of the pattern-based prognostic methodology

In what follows, the details of each step are presented.

**Step1: Data preparation**

For data preparation, we consider a set of observations collected from a group of $M$ systems at every operational cycle. The systems are working under different specific conditions. Each observation comprises the identity of the system, the time, and a set of covariates reflecting the operating conditions and the condition indicators. These covariates may be time-dependent, that is changing with the time, or time independent. Each system is observed periodically until the failure occurs (Run-To-Failure data). The time just before the system fails will be considered as the time to failure (*TTF*). The corresponding observation is called a failure one, while all the others are called normal observations. In Table 4.1, $O(i, t_k, Z_{i,t_k})$ represents the observation collected from the $i^{th}$ system $(i = 1,2, \dots M)$ at time $t_k$ $(t_k = 1,2, \dots t_{Fi})$, where $t_{Fi}$ is the failure time, and $Z_{i,t_k}$ is a vector of covariates' values at time $t_k$. The highlighted gray observations $O(i, t_{Fi}, Z_{i,t_{F_i}})$ are the failure observation of the $i^{th}$ system, and $Z_{i,t_{F_i}}$ is the corresponding vector of covariates.

The dataset listed in Table 4.1 is divided into two datasets; training dataset and updating dataset. The training dataset consist of the observations collected from a set of similar systems ($V$ systems), while the updating dataset consists of the observations that are collected from another set of similar systems ($U$ systems). This means that the total number of systems is $M = V + U$.

Table 4.1: The observations collected from $M$ system

| *Time/System* | 1 | 2 | ……. | $i$ | ……. | $M$ |
|---|---|---|---|---|---|---|
| $t_1$ | $O(1, t_1, Z_{1,t_1})$ | $O(2, t_1, Z_{2,t_1})$ | | $O(i, t_1, Z_{i,t_1})$ | | $O(M, t_1, Z_{M,t_1})$ |
| $t_2$ | $O(1, t_2, Z_{1,t_2})$ | $O(2, t_2, Z_{2,t_2})$ | | $O(i, t_2, Z_{i,t_2})$ | | $O(M, t_2, Z_{M,t_2})$ |
| . | . | . | | . | | . |
| . | . | . | ……. | . | ……. | . |
| $t_k$ | $O(1, t_k, Z_{1,t_k})$ | $O(2, t_k, Z_{2,t_k})$ | | $O(i, t_k, Z_{i,t_k})$ | | $O(M, t_k, Z_{M,t_k})$ |
| . | . | . | | . | | . |
| . | . | . | | . | | . |
| . | . | . | | . | | . |
| . | . | . | | . | | . |
| $t_{F2}$ | . | $O(2, t_{F2}, Z_{2,t_{F2}})$ | | . | | . |
| $t_{FM}$ | . | | | . | | $O(M, t_{Fi}, Z_{M,t_{Fi}})$ |
| . | . | | | . | | |
| $t_{F1}$ | $O(1, t_{F1}, Z_{1,t_{F1}})$ | | | . | | |
| . | | | | . | | |
| $t_{Fi}$ | | | | $O(i, t_{Fi}, Z_{i,t_{Fi}})$ | | |
| . | | | | | | |

***Training dataset:*** The training dataset is drawn from the dataset listed in Table 4.1. The training observations collected from $V$ systems are divided into two different categories (classes); *SL* systems and *LL* systems. The systems that failed before a certain time $t_S$, specified and decided by the maintenance personnel, are called *SL* systems and all their corresponding observations are *SL* ones, and the ones that failed after that time are called *LL* systems. The time $t_S$ is set to the *MTTF*, which can be obtained by the maintenance personnel, by using the historical lifetime data. Accordingly, $t_S$ is the separation time between the two classes (*SL* and *LL*). In Table 4.2, the training observations are classified as either $\Omega^{SL}$or $\Omega^{LL}$, where $\Omega^{SL}$and $\Omega^{LL}$ are the sets of *SL* and *LL* systems' observations, respectively. The highlighted entries in the table represent the set of failure observations.

The *SL* observations are listed in the first raw of Table 4.2 and the *LL* observations are listed in the second raw. It means that for the $v^{th}$ system, the normal observation $O(v, t_k \leq t_{Fv}, Z_{v,t_k \leq t_{Fv}})$ and the failure observation $O(v, t_{Fv} \leq t_s, Z_{v,t_{Fv} \leq t_s})$, are *SL* observations. The normal observations $O(v, t_k \leq t_s, Z_{v,t_k \leq t_s})$ and $O(v, t_k > t_s, Z_{v,t_k > t_s})$, and the failure observation $O(v, t_{Fv} > t_s, Z_{v,t_{Fv} > t_s})$, are considered to be *LL* ones. Figure 4-2 illustrates how those observations are classified according to the failure time of the corresponding system. In this figure, the $v^{th}$ system in the training dataset may fail before or after the time $t_s$. If the system fails before this time, it means that it is *SL* system, and if it fails after that time, it is *LL* one. The collected observations are plotted as red points if the system is *SL*, and in green if the system is *LL*. The small red rectangle represents the failure observation of the $v^{th}$ *SL* system, while the green rectangle represents the failure observation of the $v^{th}$ *LL* system.

The historical lifetime data of the systems are used to estimate the baseline survival curve in step 2, as shown in Figure 4-1. In step 3 as depicted in the figure, the condition monitoring covariates are employed by the two-class LAD classifier in order to generate the patterns that are discriminating between the set of *SL* observations and the set of *LL* observations.

Table 4.2: Representation of the training observations collected from $V$ systems

| *Class* | *Failure Time* | *Observations* |
|---|---|---|
| *Short Life systems* $\Omega^{SL}$ | $t_{Fv} \leq t_s$ | $O(v, t_k \leq t_{Fv}, Z_{v,t_k \leq t_{Fv}})$<br>$O(v, t_{Fv} \leq t_s, Z_{v,t_{Fv} \leq t_s})$ |
| *Long Life systems* $\Omega^{LL}$ | $t_{Fv} > t_s$ | $O(v, t_k \leq t_s, Z_{v,t_k \leq t_s})$<br>$O(v, t_k > t_s, Z_{v,t_k > t_s})$<br>$O(v, t_{Fv} > t_s, Z_{v,t_{Fv} > t_s})$ |

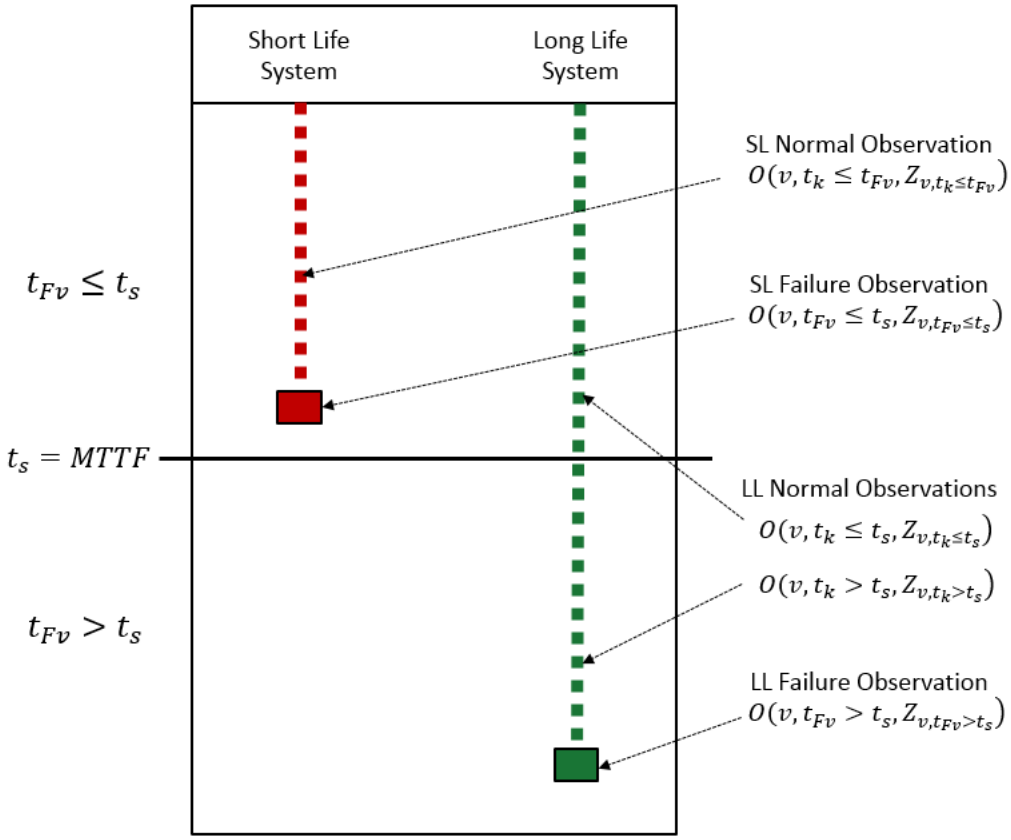


Figure 4-2: *SL* and *LL* observations in the training dataset

***Updating dataset:*** The updating dataset is drawn from Table 4.1. It is a set of observations collected from another group of systems $U$ which are not in the training dataset, in order to validate the proposed methodology. The operating conditions, the condition indicators, and the time of each observation, are considered. The observations collected from each system are used to update its survival curve, instantaneously. The survival curve of the $u^{th}$ system is updated at each time a new updating observation $O(u, t_k, Z_{u,t_k})$ is available, according to the patterns covering that observation, where $u = 1,2, \dots U$.

**Step 2: Baseline survival curve estimation using KM estimator**

In this step, the survival curve, also called survival probability function, is estimated by KM estimation technique using the lifetime data of all the *SL* and *LL* systems, in the training dataset. The baseline survival curve, is estimated using equation (4.1) as in [16]:

$$S_b(t) = \prod_{t_{Fv} \le t} [1 - \frac{d_{t_{Fv}}}{Y_{t_{Fv}}}] \tag{4.1}$$

where $d_{t_{Fv}}$ is the number of systems that failed at the time $t_{Fv}$, and $Y_{t_{Fv}}$ is the number of systems which are at risk at that time. The baseline survival function $S_b(t)$ considers all the failure observations of the two classes (*SL* and *LL* system). It does not reflect the effect of the operating conditions on the reliability of the monitored system. It reflects only the effect of the working time on the system's health state. This is one of the limitations of using the baseline alone when estimating the RUL in CBM prognostics. In the next step, LAD is used to reflect the effect of the operating conditions, by generating a set of *SL* and *LL* patterns from the condition monitoring data, in the training dataset.

**Step 3: Pattern generation using the two-class LAD classifier**

A specific characteristic of LAD is the generation of patterns which represent the hidden knowledge in the training dataset. These patterns are hidden natural rules that define the interactions between the covariates for *SL* and *LL* systems separately. Since LAD is not a statistical-based approach, there is no need to satisfy any statistical assumptions. As a result, the pattern generation is carried out even if the covariates are highly correlated or time-dependent.

The LAD classifier is used in this step to generate the *SL* and *LL* patterns that differentiate between the *SL* and the *LL* systems, in the training dataset. Given only the covariates of the observations that are found in Table 4.2, and the class labels (*SL* and *LL*), in the training dataset, the two-class LAD classifier is used to generate the set $P_{Gen}$ of *SL* and *LL* patterns. Each generated pattern $p_g$ in the set $P_{Gen}$ (i.e $p_g \in P_{Gen}$), covers only a subset of observations in the training dataset.

**Step 4: Pattern selection using an SCP formulation**

After generating the set $P_{Gen}$ of *SL* and *LL* patterns, we formulate the pattern selection as an SCP, in order to select the most significant patterns and to remove the redundant patterns. The most significant patterns are covering the largest number of observations. They are called strong patterns. This SCP guarantees that each observation in the training dataset is at least covered by one of the selected patterns.

The SCP is NP complete [28], and an exact solution for the large size problems could not be found in a reasonable time. Therefore, in our proposed methodology, the greedy algorithm developed in [28] is applied. The feasible solution of the SCP is obtained after a number of iterations. In the beginning an empty set $P_{Sel}$ is assigned to the set of selected patterns. The algorithm is initialized and the pattern that covers the maximum number of observations in the training observations is added to the set $P_{Sel}$. At each consecutive step of this algorithm, another pattern that covers the maximum number of the observations that are currently uncovered by the previously selected patterns, is added to the set of selected patterns. The algorithm stops when all the training observations are covered by the patterns that are included in the set $P_{Sel}$. The set of selected patterns $P_{Sel}$ is a subset of the generated patterns ($P_{Sel} \subseteq P_{Gen}$). It is the minimal subset of patterns such that every observation in the training dataset is covered at least once.

By ignoring the redundant patterns and considering only the selected ones, this leads to build more stable and robust classifier model, in terms of being able to classify both *SL* and *LL* observations, accurately. Consequently, this results in developing more accurate prognostic methodology for predicting the RUL, as it will be shown in Section 4.5 of the paper.

Each selected pattern has a normalized weight that reflects the ratio between the number of covered observations by this pattern to the total coverage of all the patterns in the same class. The normalized weight of the selected pattern $p_s$ ($s = 1,2, \dots, |P_{Sel}|$), which takes values between 0 and 1, is defined and calculated as:

$$W_{p_s} = \frac{cov(p_s)}{\sum_{q=1}^{Q} cov(p_q)} \tag{4.2}$$

where $cov(p_s)$ is the set of observations that are covered by the selected pattern $p_s$, and $Q$ is the number of selected patterns for the corresponding class. As will be shown later, the pattern selection procedure as well as the weights of the selected pattern, have a significant effect on the accuracy of the RUL estimation.

**Step 5: Survival curve estimation for the selected patterns using KM estimator**

In this step, the survival curve of each selected pattern is estimated by considering only the failure observations of the systems which are covered by that pattern. Accordingly, the inputs to KM estimation technique are the failure times of the observations that are covered by that pattern. The survival curve of the selected pattern $p_s$, is estimated using equation (4.3) as:

$$S_{p_s}(t) = \prod_{\substack{t_{Fv} \leq t \\ \forall\, O(v, t_{Fv}, Z_{v,Fv}) \in cov(p_s)}} \left[1 - \frac{d_{t_{Fv}}}{Y_{t_{Fv}}}\right] \tag{4.3}$$

Each pattern's survival curve represents the complement of a non-parametric cumulative distribution function (CDF), for the subset of failure observations that are covered by that pattern.

Each subset has shown specific hidden interactions between the operating conditions and the condition indicators. As such, the failure times are not any more the only measure that characterize the system reliability, but also the system's conditions are reflected in the set of patterns' survival curves that are represented in equation (4.3). After the execution of this step, the training phase of the methodology is terminated and the knowledge is extracted from the training dataset in the form of a set of survival curves; the baseline curve and the patterns' survival curves.

**Step 6: Updating the survival curve of the monitored system**

This is the beginning of the updating phase of the methodology. The estimated baseline survival curve obtained in step 2 and the survival curves for the *SL* and *LL* patterns obtained in step 5, are used to update the survival curve of each of the systems in the updating dataset. Given the observations collected from each system, LAD is used as a diagnostic decision model in order to extract the patterns that cover each observation and to update the system's survival curve. The updated survival curve not only reflects the system's age but also reflects the effect of the operating conditions on the system's health state.

Initially, the survival curve is updated by averaging the baseline survival curve and the survival curves of the patterns that are covering the most recent observation. The set of covering patterns is denoted by $P_{Cov}$. It is a subset of the set of the selected patterns (i.e. $P_{Cov} \subset P_{Sel}$). The following updating observations are then used to update the survival curve by averaging the survival curves of the patterns that cover each observation, and the former updated survival curve, simultaneously. In this paper, based on the most recent collected observation, one of the following three updating formulas is used to update the survival curve of the $u^{th}$ monitored system at time $t_k$.

| | | |
|---|---|---|
| ***Formula 1*** | $$S_{O(u,t_k,Z_{u,t_k})}(t) = \begin{cases} [\sum_{s=1}^{n} S_{p_s}(t) + S_b(t)\,]/(n+1) & \forall\ k=1 \\ [\sum_{s=1}^{n} S_{p_s}(t) + S_f(t)\,]/(n+1) & \forall\ k \geq 2 \end{cases}$$ | (4.4) |
| ***Formula 2*** | $$S_{O(u,t_k,Z_{u,t_k})}(t) = \begin{cases} [\sum_{s=1}^{n} S_{p_s}(t)/n + S_b(t)\,]/2 & \forall\ k=1 \\ [\sum_{s=1}^{n} S_{p_s}(t)/n + S_f(t)\,]/2 & \forall\ k \geq 2 \end{cases}$$ | (4.5) |
| ***Formula 3*** | $$S_{O(u,t_k,Z_{u,t_k})}(t) = \begin{cases} [\sum_{s=1}^{n} W_{p_s} S_{p_s}(t) + S_b(t)\,]/(1+\sum_{s=1}^{n} W_{p_s}) & \forall\ k=1 \\ [\sum_{s=1}^{n} W_{p_s} S_{p_s}(t) + S_f(t)\,]/(1+\sum_{s=1}^{n} W_{p_s}) & \forall\ k \geq 2 \end{cases}$$ | (4.6) |

In these updating formulas, $n$ is the number of patterns that cover the updating observation $O(u, t_k, Z_{u,t_k})$ (i.e. $n = |P_{Cov}|$), and $S_f(t)$ is the former updated survival curve obtained from the previous updating observation at time $t_{k-1}$.

*Formula 1*, which is considered in [13], has the limitation that the survival curve of the new observation is updated by assigning the same weight for each of the survival curves of the covering patterns and the baseline, although the coverage of each pattern is less than that of the baseline. We address this limitation by assigning a weight for the baseline curve that is greater than the weight of the survival curve of each pattern, as shown in *Formula 2*. Because the weights of the survival curves of the patterns are equal in *Formula 2*, we assign a different weight for each of the survival curves of the patterns in *Formula 3*. These weights reflect the coverage of each pattern. The prediction accuracies of the three formulas are compared later through a case study in Section 4.5.

**Step 7: RUL estimation**

In this step, the RUL is estimated based on the updated survival curve of the monitored system. It is important to represent the effect of the operating conditions and the condition indicators, in order to accurately predict its RUL. Let $T$ represents the time to failure which is a random variable and let $T - t_k$ represents the RUL of $T$ at time $t_k$. The RUL is random variable, its expected value called the *mean remaining useful life* (*MRUL*), is calculated using equation (4.7) as in [29]:

$$MRUL(t_k) = E[T - t_k | T > t_k] = \frac{\int_{t_k}^{\infty} S(\tau) d\tau}{S(t_k)} \tag{4.7}$$

where $S(.)$ is the survival function. The *MRUL* is calculated at each time instant a new updating observation is collected, by considering the updated survival curve of that observation. The *MRUL* is calculated by considering the vector of covariates $Z_{t_k}$ as in [30]:

$$MRUL\left(t_k, Z_{t_k}\right) = E[T - t_k | T > t_k, Z_{t_k}] = \frac{\int_{t_k}^{\infty} S(\tau, Z_\tau) d\tau}{S\left(t_k, Z_{t_k}\right)} \tag{4.8}$$

Here, the updated survival function $S_{O(u,t_k,Z_{u,t_k})}(t)$ is used instead of $S(\tau, Z_\tau)$ to give an estimate for the *MRUL* of the $u^{th}$ system at time $t_k$, since it reflects the effect of covariates on the survival probability of that system. Accordingly, the *MRUL* is given by:

$$MRUL_u\left(t_k, Z_{t_k}\right) = \frac{\int_{t_k}^{\infty} S_{O(u,t_k,Z_{u,t_k})}(\tau) d\tau}{S_{O(u,t_k,Z_{u,t_k})}(t_k)} \tag{4.9}$$

Since $P(T \geq t_k) = P(T > t_{k-1})$ for the discrete time distributions (see [31] for more details about the discrete distributions), equation (4.9) can be represented experimentally in discrete form and the *MRUL* is calculated as:

$$MRUL_u(t_k, Z_{t_k}) = E(T - t_k | T > t_{k-1}, Z_{t_k}) = \frac{\sum_{t_r=t_k}^{\infty} \Delta t_r S_{O(u,t_k,Z_{u,t_k})}(t_r)}{S_{O(u,t_k,Z_{u,t_k})}(t_{k-1})} \quad (4.10)$$

where $\Delta t_r$ is the monitoring interval which is the difference between two consecutive inspection times i.e. $\Delta t_r = t_{r+1} - t_r$. Equation (4.10) is used to estimate the RUL in the final step of this prognostic methodology.

### 4.4.3 Training and updating algorithms

Two algorithms are developed to implement the constituent steps of this prognostic methodology; algorithm 1 for the training and algorithm 2 for the updating and the RUL prediction.

Algorithm 1: Training

Input: Training dataset $\Omega^{SL} = O(v, t_k \leq t_{Fv}, Z_{v,t_k \leq t_{Fv}}), O(v, t_{Fv} \leq t_s, Z_{v,t_{Fv} \leq t_s})$ and $\Omega^{LL} = O(v, t_k \leq t_s, Z_{v,t_k \leq t_s})$, $O(v, t_k > t_s, Z_{v,t_k > t_s})$, and $O(v, t_{Fv} > t_s, Z_{v,t_{Fv} > t_s})$, where $v = \{1, \dots, V\}$.

Output: The set of generated patterns $P_{Gen}$, the set of selected patterns $P_{Sel}$, the baseline survival curve $S_b(t)$, a survival curve $S_{p_s}(t)$ for each selected pattern $p_s$, and a weight $W_{p_s}$ for each selected pattern.

1. $S_b(t)$ ← Estimation of the baseline survival curve using eq. (4.1) by considering the lifetime data $t_{Fv}$ of all systems.
2. $P_{Gen} = \{p_g, g = 1, \dots |P_{Gen}|\}\}$ ←Pattern generation using *LAD* classifier, by considering the covariates $\Omega^{SL}$ and $\Omega^{LL}$
3. $P_{Sel} = \{p_s, s = 1, \dots |P_{Sel}|\}$ ←Pattern selection procedure formulated as SCP.
4. for $s = 1$ to $|P_{Sel}|$
5.     for $v = 1$ to $V$
6.         $cov(p_s)$ ← Using LAD classifier to obtain the coverage of the pattern $p_s$, by considering $\Omega^{SL}$ and $\Omega^{LL}$.
7.     return $cov(p_s)$
8. $W_{p_s}$ ← Calculation of the weight for each selected pattern $p_s$ using eq. (4.2).
9. $S_{p_s}(t)$ ← Estimation of the survival curve for the selected pattern $p_s$ using eq. (4.3).
10. return $\{S_{p_s}(t): s = 1, \dots |P_{Sel}|\}, W_{p_s}$.

Algorithm 2: Updating the survival curve and RUL calculation for the $\boldsymbol{u^{th}}$ system

Input: $S_b(t), \{S_{p_s}(t): s = 1, \dots |P_{Sel}|\}, W_{p_s}$, and $\{O(u, t_k, Z_{u,t_k}), t_k = 1,2, \dots t_{Fu})\}$, where $u = \{1, \dots, U\}$.

Output: Updated survival curve $S_{O(u,t_k,Z_{u,t_k})}(t)$, and mean remaining useful life $MRUL_u(t_k)$.

1. for $t_k = 1$ to $t_{Fu}$
2.     $P_{Cov}$ ← Using LAD diagnostic model to obtain the set of covering pattern, by considering $Z_{u,t_k}$.
3.     $S_{O(u,t_k,Z_{u,t_k})}(t)$ ← Updating the survival curve of the $u^{th}$ system using eq. (4.4) or eq. (4.5), or eq. (4.6).
4.     $MRUL_u(t_k)$ ← MRUL estimation for the $u^{th}$ system using eq. (4.10).
5. return $S_{O(u,t_k,Z_{u,t_k})}(t), MRUL_u(t_k)$.

In the next section, we use a common dataset in the field of prognostics and we compare the RUL prediction accuracy in the case of using each updating formula; *Formula 1*, *2* and *3*, through one of the common tests; Friedman test. Moreover, the proposed prognostic methodology is compared to two common machine learning regression techniques; artificial neural networks (ANNs) and support vector regression (SVR).

## 4.5 Application To NASA Prognostic Turbofan Engine Dataset

In order to validate the proposed methodology, we use the turbofan engine dataset (C-MAPSS dataset) which is available on the website of *NASA* prognostic data repository [32]. The training dataset is collected from a set of 260 aircraft turbofan engines. Each observation in the training dataset consists of the engine identity, the age of the engine in cycles, in addition to twenty-four covariates. The first, the second, and the third covariates are three operational settings, while all the remaining covariates represent sensor measurements of condition indicators. The updating dataset consists of a set of observations collected from another set of 259 engines.

### 4.5.1 Results Analysis

#### a) Pattern Generation

The software *cbmLAD* [11] is used to generate a set of *SL* and *LL* patterns from the training observations. The engines that fail before the 206$^{th}$ cycle (which is the *MTTF*) are considered as *SL*, while those fail after that time are considered as *LL* ones. Consequently, we have 26547 *SL* observations collected from 154 *SL* systems, and 27212 *LL* observations collected from 106 *LL* systems. The *SL* and *LL* patterns are generated from the twenty-four covariates of the training observations. Accordingly, 527 *SL* patterns and 510 *LL* patterns are generated.

#### b) Pattern selection

The software *cbmLAD* applies the pattern selection procedure in order to remove the redundant patterns and to select the most significant ones. By solving an SCP, 405 patterns are selected from the 527 generated *SL* patterns (122 redundant *SL* patterns are removed), and 379 patterns are also selected from the 510 generated *LL* patterns (131 redundant *LL* patterns are removed). The survival curves for a sample of the selected *SL* and *LL* patterns are shown in Figure 3. The survival curves of the generated patterns are plotted in the figure in different colors, while the baseline survival curve is plotted in black. The curves that are below the baseline belong to the *SL* patterns, while those above the baseline belong to the *LL* patterns. The *SL* survival patterns represent the specific characteristics for the short life systems, while the *LL* patterns represent the specific characteristics for those that have long lives.

### c) Survival curve updating

To clarify how the survival curve of one of the aircraft turbofan engines is updated, we exploit the survival functions of the patterns that are plotted in Figure 4-3. The figure illustrates also the procedure when using *LAD Formula 1* in updating the survival curve of the first engine in the updating dataset. Given the first observation collected from that engine, the set of covering patterns $P_{Cov} = \{p_{13}^{-}, p_{14}^{-}, p_{17}^{+}, p_{26}^{-}\}$, is obtained by using the two-class LAD diagnostic model. As shown in the figure, the pattern $p_{17}^{+}$ is *SL* pattern, while the patterns $p_{13}^{-}, p_{14}^{-}$, and $p_{26}^{-}$ are *LL* patterns. The updated survival curve for that observation is the dotted curve in blue. After collecting the second observation, the covering *SL* patterns for that observation are $p_{18}^{+}, p_{20}^{+}, p_{23}^{+}, p_{27}^{+}$, and $p_{33}^{+}$, while the covering *LL* patterns are $p_{2}^{-}, p_{3}^{-}$, and $p_{5}^{-}$. The resulted updated curve for that observation is the dotted curve in black.

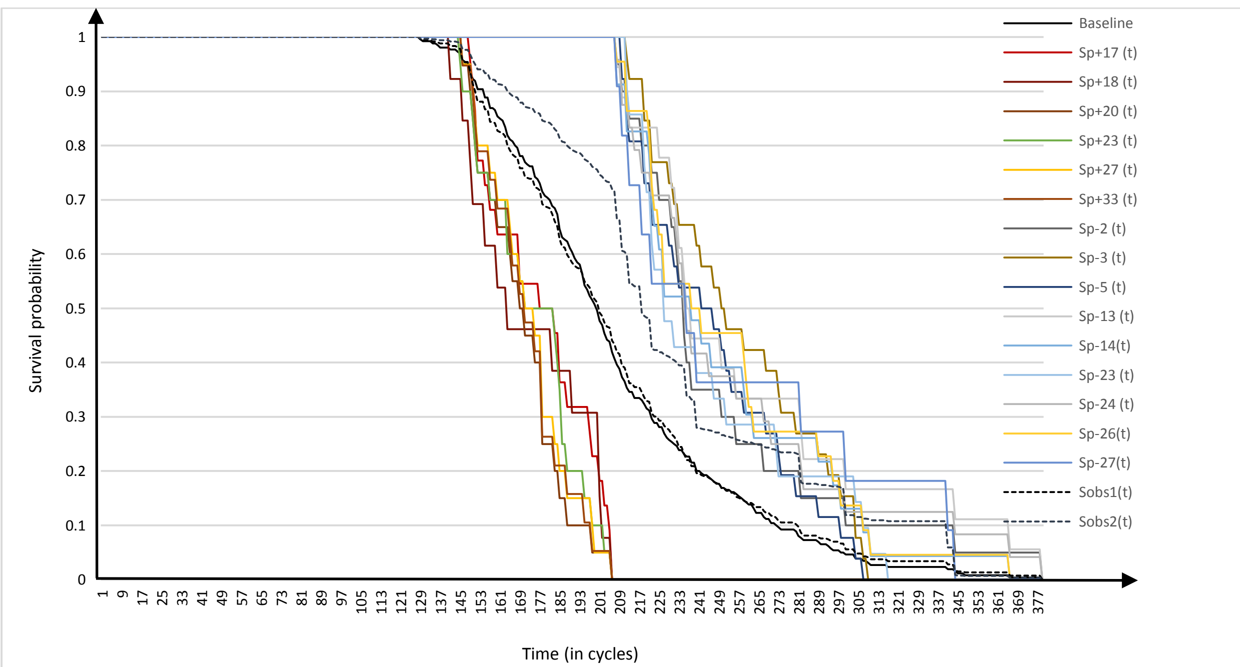


Figure 4-3: The baseline survival curve, the survival curves for a sample of the selected *SL* and *LL* patterns, and the updating procedure after collecting the first two observations from the first engine, using *LAD Formula 1*.

**d) RUL estimation**

Since each engine in the updating dataset has a different failure history, it is important to have a good accuracy measure to validate the prognostic methodology. We use the *root mean squared error* (*RMSE*) as an error measurement criterion. The *RMSE* of the *MRUL* estimation for the $u^{th}$ engine is calculated as:

$$RMSE(u) = \sqrt{\sum_{t_k=1}^{t_{Fu}} [actual\ RUL(t_k) - estimated\ MRUL(t_k)]^2 / N_{Fu}} \qquad (4.11)$$

where $N_{Fu}$ is the actual number of operational cycles until the failure of the $u^{th}$ engine, where $u = 1, 2, \ldots U$. The *RMSE* is calculated for the 259 engines using the three different updating formulas; *LAD Formula 1*, *LAD Formula 2*, and *LAD Formula 3*. The *RMSE* of the estimated *MRUL* for a sample of ten engines are listed in Table 4.3. The underlined bold results (the smallest *RMSE*) indicate that the corresponding formula gives the highest performance. It can be noticed from this sample that, in general, *LAD Formula 2* and *LAD Formula 3* give a better results than that are given by *LAD Formula 1*. However, the results listed in Table 4.3 are not necessarily the best way to compare between the three formulas. There is a need to statistically compare between the three prognostic formulas, based on the calculated *RMSE* values of all engines.

Based on the results of *MRUL* estimation for the 259 engines, we need to properly determine whether the accuracy of *LAD Formula 2* and *LAD Formula 3* significantly outperform that of *LAD Formula 1*, or not. That is, the question that needs an answer is: Whether any updating formula ranks consistently and significantly higher or lower than the others?

Table 4.3: The results for *LAD Formula 1*, *LAD Formula 2*, and *LAD Formula 3*, for a sample of ten engines

| System Identity | Time To Failure (TTF) | RMSE using *LAD Formula 1* | RMSE using *LAD Formula 2* | RMSE using *LAD Formula 3* |
|---|---|---|---|---|
| 6 | 218 | 31.87956 | 28.08579 | 26.85231 |
| 55 | 190 | 20.07931 | 17.82797 | 17.28543 |
| 85 | 263 | 39.12543 | 34.23242 | 35.27549 |
| 117 | 172 | 18.40472 | 16.33767 | 15.95682 |
| 126 | 163 | 15.18374 | 16.79543 | 15.04567 |
| 141 | 168 | 27.62458 | 26.85612 | 24.56912 |
| 162 | 147 | 20.56892 | 18.85341 | 18. 45628 |
| 175 | 194 | 28.53627 | 25.26347 | 25.78456 |
| 200 | 171 | 23.53687 | 22.38421 | 21.48267 |
| 247 | 202 | 27.38167 | 26.57846 | 25.18547 |

We need a suitable test to report the comparison of all formulas and to consider the difference in the failure history of each engine in the updating dataset. In what follows, we use Friedman test to rank the results obtained from the 259 engines, by using the three updating formulas.

### 4.5.2 Accuracy of the MRUL Calculations

**a) Comparing *LAD Formula 1*, *LAD Formula 2*, and *LAD Formula 3***

Friedman test is a non-parametric statistical test, used to compare different values of population means that are evaluated under different levels of the study's factors [33]. In this study, the first factor is the difference in the failure histories of the engines, and the second factor is the applied updating formula that affects the magnitude of the calculated *RMSE* value.

In the Friedman test of this work, two phases are followed. In the first phase, we test the hypothesis that all the applied updating formulas are statistically equivalent or not. In the second phase, a set of post tests is carried out, in order to report the formula that gives the highest performance.

**Phase 1: Test hypotheses**

Friedman test is based on ranking; it assigns ordered ranks from the smallest to the largest, to the calculated *RMSE*, by using the three prognostic formulas for each engine. In other words, each of the 259 engines rates those three different prognostic formulas (*LAD Formula 1*, *LAD Formula 2*, and *LAD Formula 3*). They are compared in this phase where the null hypothesis states that all of them have the same mean *RMSE*.

Thus, the null hypothesis is formulated as:

$$H_0: \mu_{LAD\ Formula\ 1} = \mu_{LAD\ Formula\ 2} = \mu_{LAD\ Formula\ 3}$$

and the alternative hypothesis is formulated as:

$$H_a: \text{Not all } RMSE \text{ means } (\mu_{LAD\ Formula\ 1}, \mu_{LAD\ Formula\ 2}, \mu_{LAD\ Formula\ 3}) \text{ are equal}$$

To reject or to accept this null hypothesis, the test statistic given by equation (4.12), is applied as in [33]:

$$F_r = \frac{12}{Uk(k+1)} \sum_{j=1}^{k} R_j^2 - 3U(k+1) \tag{4.12}$$

where $R_j$ is the rank sum for the $Formula\ j$ $(j = 1, 2, 3)$, and $U$ is the number of engines in the updating dataset. The values of the rank sums for the three formulas, without applying the pattern selection procedure, are listed in the second columns of Table 4.4. The mean rank of each of the three formulas for the calculated 259 *RMSE* values is listed in the third column of the table. The calculated significance level (*p-value*) for this test statistic is given by:

$$p - value = P(\chi^2_{k-1} \geq F_r) = P(\chi^2_{3-1} \geq 137.7683) = 0$$

and the declared test significant level is calculated as: $\chi^2_{k-1,\alpha} = \chi^2_{2,0.05} = 5.9914$, where $\chi^2_{k-1}$ is the Chi Square test statistic [34]. The calculated *p-value* is too small, and the test is said to be significant, and accordingly the null hypothesis $H_0$ is rejected. Rejecting $H_0$ means that some prognostic formulas tend to have larger or smaller *RMSE* means than the others.

**Phase 2: Pairwise comparisons**

After carrying out the first phase of Friedman test, there is no indication as to which formula has better performance than the others, although we are confident that one of the three updating formulas outperforms the other two. Thus, multiple post-hoc comparisons (pairwise Friedman test), are performed, in order to find the best prognostic formula, that is the formula that leads to a significantly lower *RMSE*.

In pairwise Friedman test, we select the formula that has the largest mean rank, and compare it with those of the lower mean ranks. As depicted from the third column of Table 4.4, *LAD Formula 1* has the largest mean rank. Consequently, the pairwise Friedman test is performed to compare the *RMSE* estimated using *LAD Formula 1*, and those estimated using *LAD Formula 2* and *LAD Formula 3*.

Two hypotheses for each comparison are formulated, and a decision is made regarding the rejection or acceptance of the null hypothesis. In the comparison of *LAD Formula 1* with $Formula\ j$ $(j = 2, 3)$, the null and the alternatives hypotheses are formulated as:

$$H_0: \mu_{LAD\ Formula\ 1} = \mu_{LAD\ Formula\ j}$$
$$H_a: \mu_{LAD\ Formula\ 1} > \mu_{LAD\ Formula\ j}$$

First, we compute the absolute difference between the rank sums $\left|R_{LAD\ Formula j} - R_{LAD\ Formula\ 1}\right|$, and if this difference exceeds the post-hoc value $d_{\alpha_F}\sqrt{Uk(k+1)/6}$, the null hypothesis is rejected. The value $d_{\alpha_F}$ is the $100(1 - \alpha_F)^{th}$ of the standard normal distribution, and $\alpha_F$ is the family-wise significance level [35].

The results of Friedman test, when the three formulas are applied, are presented briefly in Table 4.4. In this table, we use the set of generated patterns in the prognostic methodology, without removing the redundant ones. It is seen in the fourth column of Table 4.4 that *LAD Formula 3* outperforms both of *LAD Formula 1* and *LAD Formula 2*, through the largest rank difference between *LAD Formula 3* and *LAD Formula 1*, which is 258. This may be attributed to the fact that the weights assigned to the survival curves in the updating step of the methodology, have a significant effect on the estimated RUL. Actually, the assigned weight reflects the coverage of each pattern, and consequently reflects its importance.

Table 4.4: Friedman test without considering pattern selection in the proposed LAD methodology

| | Phase 1 | | Phase 2 | | | |
|---|---|---|---|---|---|---|
| Updating LAD Formula | Rank sum $R_j$ | Mean rank | $\lvert R_j - R_{LAD\ Formula\ 1}\rvert$ | *Is* $\lvert R_{LAD\ Formula\ j} - R_{LAD\ Formula\ 1}\rvert > d_{\alpha_F}\sqrt{Uk(k+1)/6}$? | Acceptance of the null hypothesis? | Significant difference? |
| *LAD Formula 2* | 478 | 1.846 | 189 | yes | No | yes |
| *LAD Formula 3* | 409 | 1.579 | 258 | yes | No | yes |
| *LAD Formula 1* | 667 | 2.575 | | | | |
| Test Statistic $F_r$ | 137.768 | | | | | |
| $\chi^2_{k-1,\alpha}$ | 5.9914 | | | | | |
| $p$-value | 0 | | | | | |
| $\alpha_F$ | 0.025 | | | | | |
| $d_{\alpha_F}$ | 1.96 | | | | | |
| $d_{\alpha_F}\sqrt{Uk(k+1)/6}$ | 44.609 | | | | | |

As stated previously, the objective of the pattern selection procedure is to remove the redundant patterns from the set of generated patterns. The selected patterns have a larger coverage than those of the original generated patterns. The pattern selection removes the redundant patterns, and consequently assigns larger weights for the selected patterns.

The results of Friedman test when the selected patterns are considered in the updating formulas are shown in Table 4.5. The results obtained show that *LAD Formula 3* still outperforms *LAD Formula 1* and *LAD Formula 2*. The largest rank difference is the one between *LAD Formula 3* and *LAD Formula 1* (it is 280 as shown in the fourth column of Table 4.5). It is noticed that the pattern selection procedure has an effect on all the updating formulas. The effect of using the selected patterns is noticed in this table by comparing the average ranking that are changed between the *LAD Formula 1* and *LAD Formula 2* in favor of the latter. The difference in rank increases from 189 as shown in Table 4.4 to 206 as shown in Table 4.5. Moreover, the pattern selection has also a significant effect on the *LAD Formula 3*, since the new weights of the selected patterns are affecting the results; the highest rank difference is increased to 280 instead of 258 as in Table 4.4. From these two tables, it is clear that the results obtained by using *LAD Formula 2* and *LAD Formula 3* have significant differences when compared to those obtained by using *LAD Formula 1*. It is concluded that the selected patterns and the weights of the patterns have a significant effect on the results.

Table 4.5: Friedman test when considering the selected patterns

| | Phase 1 | | Phase 2 | | | |
|---|---|---|---|---|---|---|
| Updating LAD Formula | Rank sum $R_j$ | Mean rank | $\lvert R_j - R_{LAD\ Formula\ 1}\rvert$ | *Is* $\lvert R_{LAD\ Formula\ j} - R_{LAD\ Formula\ 1}\rvert > d_{\alpha_F}\sqrt{Uk(k+1)/6}$? | Acceptance of the null hypothesis? | Significant difference? |
| *LAD Formula 2* | 474 | 1.830 | 206 | Yes | No | Yes |
| *LAD Formula 3* | 400 | 1.544 | 280 | Yes | No | Yes |
| *LAD Formula 1* | 680 | 2.626 | | | | |
| Test Statistic $F_r$ | 162.564 | | | | | |
| $\chi^2_{k-1,\alpha}$ | 5.9914 | | | | | |
| $p$-value | 0 | | | | | |
| $\alpha_F$ | 0.025 | | | | | |
| $d_{\alpha_F}$ | 1.96 | | | | | |
| $d_{\alpha_F}\sqrt{Uk(k+1)/6}$ | 44.609 | | | | | |

## b) Comparing the proposed methodology to the methodology introduced in [13]

The Friedman test is performed to compare the proposed prognostic methodology and the methodology proposed in [13]. The test allows evaluating the effect of using all the condition monitoring data versus using only failure data. In this test, we compare the results obtained by using the two updating formula in both methodologies. The results of comparing the formulas in both methodologies are listed in Table 4.6 and Table 4.7. It is shown that the proposed methodology outperforms the methodology proposed in [13].

Table 4.6: Comparison between the obtained results using the first two formulas in both methodologies

| | Phase 1 | | Phase 2 | | | |
|---|---|---|---|---|---|---|
| Prognostic Methodology | Rank sum | Mean rank | $\lvert R_{LAD\ Formula\ 1} - R_{LAD\ Model\ 1}\rvert$ | *Is* $\lvert R_{LAD\ Formula\ 1} - R_{LAD\ Model\ 1}\rvert > d_{\alpha_F}\sqrt{Uk(k+1)/6}$? | Acceptance of the null hypothesis? | Significant difference? |
| *LAD Formula 1* in our methodology | 300 | 1.158 | 177 | Yes | No | Yes |
| *LAD Model 1* in [13] | 477 | 1.842 | | | | |
| Test Statistic $F_r$ | 120.96 | | | | | |
| $\chi^2_{k-1,\alpha}$ | 3.84 | | | | | |
| $p$-value | 0 | | | | | |
| $\alpha_F$ | 0.05 | | | | | |
| $d_{\alpha_F}$ | 1.65 | | | | | |
| $d_{\alpha_F}\sqrt{Uk(k+1)/6}$ | 26.47 | | | | | |

Table 4.7: Comparison between the obtained results using the second two formulas in both methodologies

| | Phase 1 | | Phase 2 | | | |
|---|---|---|---|---|---|---|
| Prognostic Methodology | Rank sum | Mean rank | $\lvert R_{LAD\ Formula\ 2} - R_{LAD\ Model\ 2} \rvert$ | *Is* $\lvert R_{LAD\ Formula\ 2} - R_{LAD\ Model\ 2} \rvert > d_{\alpha_F}\sqrt{Uk(k+1)/6}$? | Acceptance of the null hypothesis? | Significant difference? |
| *LAD Formula 2* in our methodology | 298 | 1.151 | 181 | Yes | No | Yes |
| *LAD Model 2* in [13] | 479 | 1.849 | | | | |
| Test Statistic $F_r$ | 126.49 | | | | | |
| $\chi^2_{k-1,\alpha}$ | 3.84 | | | | | |
| *p*-value | 0 | | | | | |
| $\alpha_F$ | 0.05 | | | | | |
| $d_{\alpha_F}$ | 1.65 | | | | | |
| $d_{\alpha_F}\sqrt{Uk(k+1)/6}$ | 26.47 | | | | | |

### c) Comparing the proposed methodology to artificial neural networks (ANN) and support vector regression (SVR)

The proposed methodology is compared to the most common machine learning regression techniques: artificial neural networks (*ANNs*) [36, 37] and support vector regression (*SVR*) [38]. They are used commonly as regression techniques to predict the remaining useful life in the field of CBM prognostics [10, 39, 40]. The variable of interest (dependent variable) in the two regression techniques is numerical rather than categorical. The task in this case is to find a relationship between the dependent variable and a group of independent variables. In this study, the remaining useful life is the dependent variable, while the independent variables are the set of covariates that represent the operating conditions and the condition indicators. In what follows, each technique is presented briefly along with the experimentations that are carried out in this paper.

**Artificial Neural Networks.** ANNs have given many desirable characteristics that are not present in many machine learning regression techniques. These include generalization ability, learning ability, and adaptability [36]. They are used to represent the nonlinear relationship between the covariates and the RUL in order to get more accurate and precise predictions. The *ANN* model was trained using the turbofan engine dataset. The architecture of the trained model has four layers of neurons; one input layer, two hidden layers, and one output layer. The *ANN* is trained using the collected observations from the 260 engines. The inputs to the *ANN* model are the age and condition monitoring data, while the desired output is the actual RUL (the difference between the equipment failure time and its current age). Each input neuron represents a separate covariate and accordingly the number of the neurons in the input layer is equal to the number of covariates in the dataset (i.e. 24 neurons in the input layer). After a number of experimentation, the optimal number of nodes for each of the hidden layers is set to seven in this paper. The weights of the *ANN* are adjusted based on gradient-descent method [38], based on the difference between the actual and the estimated RUL, in the training phase. After the training phase, the network can recognize the correlation and relates the observations to the actual RUL. Given the collected observations from the 259 engines in the updating dataset, the trained model is used to predict the RUL for each. The *RMSE* for each engine is calculated as in equation (4.11).

**Support Vector Regression.** Support vector regression is regression technique based on the concept of support vector machines (*SVM*) that was developed by Vapnik [41]. Recently, *SVR* becomes competitive with the best available regression techniques, and it has now evolved into an active area of research. A comprehensive tutorial on *SVR* has been published by [42]. The *SVR* has two main important properties. First, it has better generalization ability than the other competitive techniques due to choosing the maximal margin hyperplane, thus minimizing the risk of over-fitting [38]. Second, it supports an efficient learning for highly nonlinear functions by applying the kernel trick [41]. According to these two properties, it is expected that *SVR* will give better prediction results than those given by *ANNs*. The objective is to estimate the parameters of certain function which give the best fit of the covariates of the training data. Such function approximates all pairs while maintaining the differences between estimated values and real values under a certain precision [40]. The kernel trick is applied to transform the input space (the set of covariates) to high dimensional feature space, and the *SVR* in this space becomes a nonlinear function in the original covariates. There are several kernels; choosing one kernel is an application-dependent (it depends on the task at hand). The commonly used family of kernels are; polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel. In this paper, we tested and applied the RBF kernel, and it gives the best results.

**Implementation of ANN and SVR models and performance comparison with LAD formulas**

The algorithms for *ANN* and *SVR* models are implemented in the publicly available *Weka* software package [38, 43]. The results of comparing the performance of all prognostic models (*LAD Formula 1*, *LAD Formula 2*, *LAD Formula 3*, *ANN* and *SVR*) are depicted in Table 4.8 and Table 4.9. In Table 4.8, we did not use the pattern selection procedure in the three *LAD* formulas. Table 4.9 shows the results when the pattern selection procedure is used. It is noticed from Table 8 that the *ANN* has the largest mean rank. Thus, we compare the performance of the *ANN* to all the remaining prognostic models. It is seen in the fourth column of Table 4.8 that *LAD Formula 2* and *LAD Formula 3* outperform both of *ANN* and *SVR*. This can be clarified by inspecting the rank differences between the *ANN* and the two models; *LAD Formula 2* and *LAD Formula 3,* which are

265 and 368, respectively. It is also noticed that these differences are larger than the difference between the *ANN* and the *SVR*, which is 92. As a result of using the pattern selection procedure, these differences increase as shown in the fourth column of Table 4.9, in particular for *LAD Formula 3*.

The final decision after the comparisons listed in Table 4.8 and Table 4.9 can be summarized as follow: there is a significant difference between the performance of *LAD Formula 2 and LAD Formula 3*, and that of *ANN*. Moreover, *LAD Formula 2 and LAD Formula 3* outperform also the *SVR*. There is no significant difference between the performance when using *ANN*, and that of using *LAD Formula 1*. There is a significant difference between the performance of *ANN* and *SVR.*

Table 4.8: Friedman test for the ANN, the SVR, and the proposed LAD methodology without considering pattern selection

| | Phase 1 | | Phase 2 | | | |
|---|---|---|---|---|---|---|
| Prognostic Model | Rank sum $R_j$ | Mean rank | $\lvert R_j - R_{ANN}\rvert$ | $\lvert R_{Model\ j} - R_{ANN}\rvert > d_{\alpha_F}\sqrt{Uk(k+1)/6}$ ? | Acceptance of the null hypothesis? | Significant difference? |
| LAD Model 1 | 870 | 3.3591 | 65 | No | Yes | No |
| LAD Model 2 | 670 | 2.5869 | 265 | Yes | No | Yes |
| LAD Model 3 | 567 | 2.1892 | **368** | Yes | No | Yes |
| SVR Model | 843 | 3.2548 | 92 | Yes | No | Yes |
| ANN Model | 935 | 3.6100 | | | | |
| Test Statistic $F_r$ | 144.4293 | | | | | |
| $\chi^2_{k-1,\alpha}$ | 9.488 | | | | | |
| $p$-value | 0 | | | | | |
| $\alpha_F$ | 0.0125 | | | | | |
| $d_{\alpha_F}$ | 2.241 | | | | | |
| $d_{\alpha_F}\sqrt{Uk(k+1)/6}$ | 80.6449 | | | | | |

Table 4.9: Friedman test for the ANN, the SVR, and the proposed LAD methodology when considering the selected patterns

| | Phase 1 | | Phase 2 | | | |
|---|---|---|---|---|---|---|
| Prognostic Model | Rank sum $R_j$ | Mean rank | $\lvert R_j - R_{ANN}\rvert$ | $\lvert R_{Model\ j} - R_{ANN}\rvert > d_{\alpha_F}\sqrt{Uk(k+1)/6}$ ? | Acceptance of the null hypothesis? | Significant difference? |
| LAD Model 1 | 884 | 3.4131 | 54 | No | Yes | No |
| LAD Model 2 | 672 | 2.5946 | 266 | Yes | No | Yes |
| LAD Model 3 | 557 | 2.1506 | **381** | Yes | No | Yes |
| SVR Model | 834 | 3.2201 | 104 | Yes | No | Yes |
| ANN Model | 938 | 3.6216 | | | | |
| Test Statistic $F_r$ | 154.5081 | | | | | |
| $\chi^2_{k-1,\alpha}$ | 9.488 | | | | | |
| $p$-value | 0 | | | | | |
| $\alpha_F$ | 0.0125 | | | | | |
| $d_{\alpha_F}$ | 2.241 | | | | | |
| $d_{\alpha_F}\sqrt{Uk(k+1)/6}$ | 80.6449 | | | | | |

## 4.6 Conclusions

In this paper, we propose a new pattern-based prognostic methodology using KM as a time-driven technique and LAD as an event-driven diagnostic method. LAD offers some important advantages, particularly when it deals with highly correlated or time varying covariates, without making any statistical assumptions. Kaplan-Meier estimator on the other hand, is non-parametric estimation method that does not need any statistical assumptions. As the estimated KM survival curve reflects only the effect of age on the systems' health state, the effect of the operating conditions of the monitored system is characterized through LAD.

The quality of the collected data affects the accuracy of the classification and consequently affects the performance of the prognostic model. Another advantage of using the LAD is its robustness to noisy and missing data, which is a big challenge in CBM prognostics.

The main objective of the methodology is to update the survival curve of the monitored system, based on the analysis of the most recent collected observation, and by finding the hidden patterns that cover this observation. Therefore, the survival curve of the monitored system is updated using the weighted sum of the survival curves of these patterns and the baseline survival curve. This updated survival curve is used to estimate the system's RUL. The proposed prognostic methodology does not set any threshold for the updated survival curve to predict the RUL.

The paper proposes two modifications for the updating formula that is used in the field of medical prognosis (*LAD Formula 1*). In the first modification, we assign a weight for the baseline survival curve that is greater than the weights of the survival curves of the patterns. In the second modification, we assign a weight for each survival curve. The assigned weight reflects the coverage of the pattern. We also consider the former updated survival curve in the updating formulas, in order to reflect the operational history of the monitored system.

It is noticed from the computational results that there is a significant difference between the performance of the two modified LAD prognostic formulas (*LAD Formula 2* and *LAD Formula 3*) and that of *LAD Formula 1*. This difference is in favor of the formers. From the obtained results, it is also concluded that the pattern selection procedure has an effect on the performance of LAD prognostic methodology because it removes the redundant patterns. The comparisons between the three prognostic formulas, using Friedman test, show that the proposed methodology yields an improved accuracy for the RUL estimation, when it considers the selected patterns and their weights. The prognostic methodology is also compared to ANN and SVR as common machine learning regression techniques. The results show that there is a significant difference between the performance of *LAD Formula 2* and *LAD Formula 3,* and that of *ANN* and *SVR.* In other words, we can conclude that *LAD Formula 3* and *LAD Formula 2* are the best, *SVR* and *LAD Formula 1* come in second, and *ANN Model* is the worst.

As a final conclusion, the proposed prognostic methodology is promising for estimating the RUL in the field of CBM prognostics since the survival curve is updated without making any prior statistical assumption. It also deals with the covariates that are highly correlated as in the case of many practical situations. The obtained results show that the proposed methodology exploits effectively the CBM data, and give accurate prediction for the RUL.

## 4.7 References

[1] R. K. Mobley, *An introduction to predictive maintenance*: Butterworth-Heinemann, 2002.

[2] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mechanical Systems and Signal Processing,* vol. 24, pp. 1161-1171, 2010.

[3] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing,* vol. 25, pp. 1803-1836, 2011.

[4] R. Kothamasu, S. H. Huang, and W. H. VerDuin, "System health monitoring and prognostics—a review of current paradigms and practices," *The International Journal of Advanced Manufacturing Technology,* vol. 28, pp. 1012-1024, 2006.

[5] A. Ghasemi, S. Yacout, and M.-S. Ouali, "Evaluating the reliability function and the mean residual life for equipment with unobservable states," *Reliability, IEEE Transactions on,* vol. 59, pp. 45-54, 2010.

[6] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing,* vol. 20, pp. 1483-1510, 2006.

[7] E. A. Elsayed, *Reliability engineering*: Wiley Publishing, 2012.

[8] A. Ghasemi, S. Yacout, and M.-S. Ouali, "Parameter estimation methods for condition-based maintenance with indirect observations," *Reliability, IEEE Transactions on*, vol. 59, pp. 426-439, 2010.

[9] G. J. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent fault diagnosis and prognosis for engineering systems*: Wiley Online Library, 2006.

[10] Z. Tian, L. Wong, and N. Safaei, "A neural network approach for remaining useful life prediction utilizing both failure and suspension histories," *Mechanical Systems and Signal Processing*, vol. 24, pp. 1542-1555, 2010.

[11] S. Yacout, D. Salamanca, and M.-A. Mortada, "Tool and method for fault detection of devices by condition based maintenance," ed: Google Patents, 2011.

[12] M.-A. Mortada, S. Yacout, and A. Lakis, "Fault diagnosis in power transformers using multi-class logical analysis of data," *Journal of Intelligent Manufacturing*, pp. 1-11, 2013.

[13] A. Ragab, M.-S. Ouali, S. Yacout, and H. Osman, "Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan–Meier estimation," *Journal of Intelligent Manufacturing*, pp. 1-16, 2014.

[14] E. Bores, P. L. Hammer, T. Ibaraki, A. Kogan, E. Mayoraz, and I. Muchnik, "An implementation of logical analysis of data," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 12, pp. 292-306, 2000.

[15] A. Bennane and S. Yacout, "LAD-CBM; new data processing tool for diagnosis and prognosis in condition-based maintenance," *Journal of Intelligent Manufacturing*, vol. 23, pp. 265-275, 2012.

[16] J. Klein and M. Moeschberger, "Survival Analysis: Techniques for Censored and Truncated Data. 1997," *Spring-Verlag, New York*, pp. 276-278.

[17] M.-A. Mortada, S. Yacout, and A. Lakis, "Diagnosis of rotor bearings using logical analysis of data," *Journal of Quality in Maintenance Engineering*, vol. 17, pp. 371-397, 2011.

[18] L. P. Kronek and A. Reddy, "Logical analysis of survival data: prognostic survival models by detecting high-degree interactions in right-censored data," *Bioinformatics*, vol. 24, pp. i248-i253, 2008.

[19] Y. Crama, P. L. Hammer, and T. Ibaraki, "Cause-effect relationships and partially defined Boolean functions," *Annals of Operations Research*, vol. 16, pp. 299-325, 1988.

[20] H. S. Ryoo and I. Y. Jang, "Milp approach to pattern generation in logical analysis of data," *Discrete Applied Mathematics*, vol. 157, pp. 749-761, 2009.

[21] T. Bonates, P. L. Hammer, and A. Kogan, "Maximum patterns in datasets," *Discrete Applied Mathematics,* vol. 156, pp. 846-861, 2008.

[22] P. L. Hammer and T. O. Bonates, "Logical analysis of data—an overview: from combinatorial optimization to medical applications," *Annals of Operations Research,* vol. 148, pp. 203-225, 2006.

[23] G. Alexe, S. Alexe, P. L. Hammer, and A. Kogan, "Comprehensive vs. comprehensible classifiers in logical analysis of data," *Discrete Applied Mathematics,* vol. 156, pp. 870-882, 2008.

[24] P. L. Hammer, A. Kogan, B. Simeone, and S. Szedmák, "Pareto-optimal patterns in logical analysis of data," *Discrete Applied Mathematics,* vol. 144, pp. 79-102, 2004.

[25] C. Guo and H. S. Ryoo, "Compact MILP models for optimal and Pareto-optimal LAD patterns," *Discrete Applied Mathematics,* 2012.

[26] J. Han, N. Kim, B. J. Yum, and M. K. Jeong, "Pattern selection approaches for the logical analysis of data considering the outliers and the coverage of a pattern," *Expert Systems with Applications,* vol. 38, pp. 13857-13862, 2011.

[27] S. Alexe, E. Blackstone, P. L. Hammer, H. Ishwaran, M. S. Lauer, and C. E. Pothier Snader, "Coronary risk prediction by logical analysis of data," *Annals of Operations Research*, vol. 119, pp. 15-42, 2003.

[28] V. Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of operations research,* vol. 4, pp. 233-235, 1979.

[29] D. Banjevic and A. Jardine, "Remaining useful life in condition based maintenance: Is it useful?," *Modelling in Industrial Maintenance and Reliability*, p. 7, 2007.

[30] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation–A review on the statistical data driven approaches," *European Journal of Operational Research,* vol. 213, pp. 1-14, 2011.

[31] M. Pintilie, *Competing risks: a practical perspective* vol. 58: John Wiley & Sons, 2006.

[32] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," 2008, pp. 1-9.

[33] W. W. Daniel, "Applied nonparametric statistics," 1990.

[34] D. C. Montgomery, *Applied Statistics and Probability for Engineers 6th edition*: Wiley, 2013.

[35] K. L. Gwet, *The Practical Guide to Statistics: Applications with Excel, R, and Calc*: Advanced Analytics Press, 2011.

[36] B. Yegnanarayana, *Artificial neural networks*: PHI Learning Pvt. Ltd., 2009.

[37] C. M. Bishop and SpringerLink, *Pattern recognition and machine learning* vol. 4: springer New York, 2006.

[38] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2011.

[39] A. Heng, A. C. C. Tan, J. Mathew, N. Montgomery, D. Banjevic, and A. K. S. Jardine, "Intelligent condition-based prediction of machinery reliability," *Mechanical Systems and Signal Processing,* vol. 23, pp. 1600-1614, 2009.

[40] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Remaining useful life estimation based on nonlinear feature reduction and support vector regression," *Engineering Applications of Artificial Intelligence,* vol. 26, pp. 1751-1760, 2013.

[41] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in neural information processing systems,* pp. 281-287, 1997.

[42] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing,* vol. 14, pp. 199-222, 2004.

[43] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann*, et al.*, "WEKA---Experiences with a Java Open-Source Project," *The Journal of Machine Learning Research,* vol. 11, pp. 2533-2541, 2010.

# CHAPTER 5 ARTICLE 3 : PROGNOSTICS OF MULTIPLE FAILURE MODES IN ROTATING MACHINERY USING LOGICAL ANALYSIS OF DATA AND CUMULATIVE INCIDENCE FUNCTIONS

Ahmed Ragab, Mohamed-Salah Ouali, Soumaya Yacout, and Hany Osman

Submitted to: JOURNAL OF MECHANICAL SYSTEMS AND SIGNAL PROCESSING

Manuscript Number: MSSP14-705

## 5.1 Abstract

This paper presents a novel methodology for multiple failure modes prognostics in rotating machinery. The proposed methodology merges a machine learning approach called Logical Analysis of Data (LAD), with a set of non-parametric cumulative incidence functions (CIFs). It considers the condition monitoring data collected from a system that experiences several competing failure modes over its life span. LAD as a non-statistical classification technique captures the actual state of the system based on the condition monitoring data. The CIF provides an estimate for the marginal probability of each failure mode in the presence of other competing failure modes. Accordingly, the assumption of the independency between the failure modes which is considered in many prognostic methods, is irrelevant in this paper. The proposed methodology is validated using vibration data collected from bearing test rigs. The results obtained are compared to those of two common machine learning prediction techniques. The comparison shows that the proposed methodology is capable of estimating accurately the RUL of an individual system, based on its collected condition monitoring data.

## 5.2 Introduction

Failure prognostics is an important decision making procedure in the condition-based maintenance (CBM) [1]. It deals with the prediction of the failure before it occurs. The task of such prediction is to determine if the failure is impending and estimate when and how this failure will occur [2]. In many industrial cases, the prognostics still relies on the experts who have significant experience about the system degradation. Recently, researches have been undertaken to develop prognostic models in order to predict the remaining useful life (RUL) of a monitored system [3]. Many of those prognostic models aim at predicting the RUL in the presence of a single failure mode, in which the historical lifetime data follow only one failure distribution with different parameters, in the presence of a set of covariates that represent the operating conditions and the condition indicators, which influence the system lifetime.

The prognostic methodology proposed in [4] uses Kaplan-Meier (KM) estimator as a non-parametric statistical method, to estimate the RUL. One of the advantages of that methodology is that it does not have any assumption about the distribution of the historical lifetime data. However, it considers the lifetime data that are drawn from only one failure mode.

The conventional prognostic methods that consider a single failure mode are inadequate to model the failure of a system subjected to multiple competing failure modes. Unfortunately, most of the multiple fault prognostic models have been theoretical and restricted to a small number of failure modes. Multiple fault prognostic models need to be both compliant and affordable to the maintenance decision makers and the technician as well. In the literature, a number of multiple fault prognostic models have already been applied in industry, although they have some theoretical limitations. They tend to concentrate on the technical merits only. A good review covering those models is found in [5].

The common method for analyzing the collected data from a set of systems that are subjected to multiple failure modes, is based on the analysis of each failure mode separately. There are two primary limitations to that method. The first limitation is that the assumption of the independency between the failure modes must be satisfied. The typical statistical survival analysis assumes that such assumption is met when there are competing failure modes, even if this is not the case in practical situations. The second limitation comes when statistical models are applied to estimate the survival curve of the system, in the presence of those competing failure modes. In such statistical models, all competing failure modes are treated as censored categories. As an example, in traditional KM estimation, the estimated survival curve for each failure mode has questionable interpretation. The estimated KM survival curve is not informative in this case as when only one failure mode is considered (the interested readers can be referred to [6] for more details). This is because KM method is mainly based on the independent censoring, which requires that those systems are censored due to other failure modes, withdrawal from the field, or lost-to-monitor at a certain time, are as likely to have a failure later as those in the risk set.

The assumption of the independence between the competing failure modes is unrealistic and mostly cannot be satisfied in the practical situations. No one can verify or prove in an explicit way that competing failure modes are independent. Three strategies that are dealing with the assumption of independence, are found in [7]. However, there is no strategy that can directly assess the independence or guarantee the correct estimate for the targeted statistical model, when the independence assumption is violated.

It is also important to reflect the effect of the operating conditions and condition indicators of the monitored system, on the failure time of each failure mode. Therefore, the big challenge in multiple fault prognostics is to evaluate the effects of each failure mode on the other competing modes, and to construct a model for estimating the RUL for the monitored systems when the covariates are highly correlated and time-varying. Thus, we need to develop prognostic models that are describing the interactions among the failure modes, and such models have to be verified and validated for varying and inter-correlated covariates.

In this paper, we propose a multiple failure modes prognostic methodology based on a supervised machine learning approach called Logical Analysis of Data (LAD), and a set of non-parametric cumulative incidence functions (CIFs). The proposed methodology consists of two phases; training and updating. In the training phase, the multiclass LAD approach is used to extract the hidden knowledge from the condition monitoring data that are collected from a set of similar systems subjected to a multitude of competing failure modes. Each observation in the condition monitoring data consists of the time and a set of covariates. The observation represents either the operational state or the failure state. The knowledge is extracted from the set of covariates in the form of relevant patterns. LAD reflects the effect of the condition monitoring data of each failure mode on the failure time of the system. A CIF curve is estimated for each failure mode, by considering the failure times of the systems that are subjected to that failure mode. In the updating phase, an observation from a newly monitored system is collected, the multi-class LAD diagnostic model is used to identify the hidden patterns which are covering that observation. The diagnostic knowledge identified from this model is then used to update the survival curve of the monitored system in order to estimate its RUL.

The effectiveness of the proposed methodology is validated using vibration data obtained from a bearing test rigs. The methodology is compared with two common machine learning prediction models; the support vector regression (SVR) and the artificial neural networks (ANNs). It is also compared with a prognostic methodology that neglects the effect of the varying failure modes on the lifetime of the monitored system.

This paper is organized in seven sections. Section 5.3 states the problem of multiple failure modes prognostics in the CBM. In Section 5.4, the theory of the CIF is presented in details. Section 5.5 presents the multi-class LAD approach, its stages, and how it is used as a classification scheme for the diagnosis of multiple faults. Section 5.6 presents the proposed multiple failure modes prognostic methodology in details. Section 5.7 presents a case study from the industry in order to validate and compare the proposed methodology to the other prediction models. Section 5.8 concludes the paper.

## 5.3 Multiple Failure Modes Prognostics

### 5.3.1 The main challenges in multiple failure modes prognostics

A big challenge in multiple failure modes prognostics is that the relationship between the RUL of the system and the covariates, is complicated and non-understandable. This is due to the competition and the interaction between the failure modes. Figure 5-1 shows how the paths of each failure mode can be interacted and have stochastic nature. The figure illustrates the competition between two failure modes. Each failure mode has one condition indicator, and the failure times of the monitored systems in each mode are determined based on a certain threshold. It is noticed from the figure that the failure times in both modes are interleaved.

Generally, this situation will be more complicated when a multitude of failure modes are considered. Therefore, it is necessary to evaluate the effects of each failure mode on the other competing modes and to build a prediction model to estimate the worst case scenario for the affected monitored system. Another challenge exists when the covariates are time-varying. Thus, the challenge is to develop prognostic models that are describing the effect of the varying operating conditions and condition indicators on the failure modes and their interactions.

### 5.3.2 The idea of the proposed methodology: Merging LAD and CIFs

The purpose of this paper is to propose a novel methodology for multiple failure modes prognostics. In this methodology, we consider the CBM data in which each system can experience one of several types of competing failure modes over its life span. The general objective is to assess the relationship of relevant covariates to the failure rate or the corresponding survival probability of any of the possible failure modes, while allowing the other competing failure modes to occur. This objective can be achieved by merging the CIF as a non-parametric time driven technique and LAD as an event driven technique.

The advantage of using the CIF curves is that the assumption of the independency of failure modes, is not necessary to be met. This is because the CIF of a certain failure mode provides an estimate for the marginal probability of that mode in the presence of the other competing modes [8]. As a result, the characteristic information of each failure mode are extracted in the form of CIF curve. The advantage of using LAD is that it is a non statistical-based classification approach that deals with the covariates that are time-varying and highly correlated. The LAD diagnostic model is used to detect novel events so that prognostic estimates can be updated appropriately. This can be achieved through updating the survival curve of the monitored system. Thereafter, the updated curve is used for the prediction of its remaining useful life.

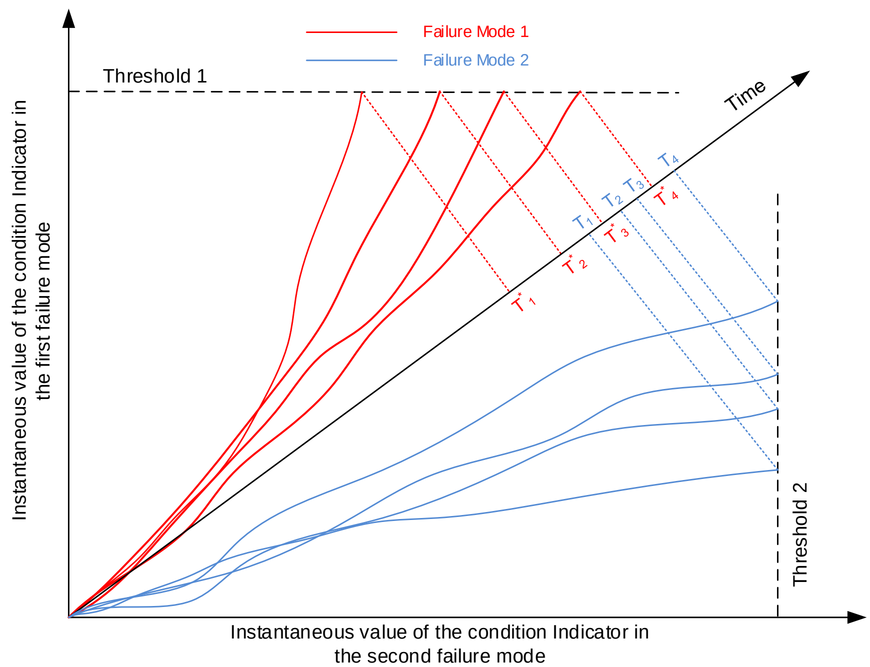


Figure 5-1: Different deterioration paths for each failure mode

In the next section, the theory of the CIF is presented in details. The LAD prognostic methodology is presented in Section 5.5 of the paper.

## 5.4 Cumulative Incidence Functions (CIFs)

The CIF is a non-parametric model that does not need any statistical assumptions to be met, as introduced in [9]. It provides an estimate for the marginal probability of a certain failure mode in the presence of the other competing failure modes, based on the collected lifetime data. The cumulative incidence model is derived from the cause-specific hazard function. This technique uses the failure times of the different failure modes to calculate the cumulative incidence rather than the survival probability [8]. Moreover, the assumption of independence between the competing failure modes is not required to be considered. In this section, the CIF is presented in details. It illustrates how the CIFs for the failure modes are estimated. It also discusses the limitation of the KM estimator when multiple failure modes are considered.

### 5.4.1 The CIF estimation for each failure mode

In the reliability analysis, it is known that the cumulative distribution function (CDF) is the probability that any failure mode occurs at or before the time $t$. Given the time to failure (TTF) which is a continuous random variable denoted by $T$, the CDF is given as presented in [10] by:

$$F(t) = Pr(T \leq t) = \int_0^t f(\tau) d\,\tau \tag{5.1}$$

The probability density function $f(\tau)$ is defined as:

$$f(t) = h(t)S(t) \tag{5.2}$$

where $h(t)$ is the hazard function, and $S(t) = 1 - F(t)$ is the survival probability (reliability) function.

The KM estimator is one of the most common non-parametric methods to estimate the survival function $S(t)$. The survival probability when considering only one failure mode is estimated using the KM estimation as presented in [11], as follows:

$$\hat{S}(t) = \prod_{t_j \leq t} [1 - \frac{d_j}{n_j}] \tag{5.3}$$

where $d_j$ is the number of systems that have experienced a failure at $t_j$ and $n_j$ is the number of systems at risk at $t_j$.

According to equation (5.2), the CDF is given by:

$$F(t) = \int_0^t h(\tau)S(\tau) d\,\tau \tag{5.4}$$

The cumulative incidence function (CIF) presented in [8], for the failure mode $i (i = 1, \ldots, C)$, is defined as follows:

$$F_i(t) = Pr(T \leq t, i) = \int_0^t f_i(\tau) d\,\tau = \int_0^t h_i(\tau)S(\tau) d\,\tau \tag{5.5}$$

where $h_i(\tau)$ is the cause-specific hazard (sub-hazard) function for the failure mode $i$. The CIF is not a proper distribution that has the properties of a distribution function, hence the term ‘sub-distribution’ is used in [12].

The cause-specific survival (sub-survival) for the failure mode $i$ is given by:

$$S_i(t) = q_i - F_i(t) \tag{5.6}$$

where $q_i = Pr(i) = \lim_{t\to\infty} F_i(t) = \frac{m_i}{M} < 1$, and $\sum_{i=1}^{C}[F_i(t) + S_i(t)] = \sum_{i=1}^{C} q_i = 1$, where $C$ is the number of failure modes. The value $m_i$ is the number of systems that failed due to the failure mode $i$ and $M$ is the total number of systems. Accordingly, the value $q_i$ physically represents the proportion of systems that failed due to the failure mode $i$.

In the case of discrete random variable $T$, the sub-hazard function for the failure mode $i$ at $t_j$ is given by $h_i(t_j) = Pr(T = t_j, i \mid T \geq t_j)$. Since $Pr(T \geq t_j) = Pr(T > t_{j-1}) = S(t_{j-1})$, it follows that:

$$h_i(t_j) = \frac{f_i(t_j)}{S(t_{j-1})} \tag{5.7}$$

where $f_i(t_j)$ is the sub-density function for the failure mode $i$ at time $t_j$. Accordingly, this sub-density function is estimated as:

$$\hat{f}_i(t_j) = \hat{h}_i(t_j)\hat{S}(t_{j-1}) \tag{5.8}$$

where $\hat{S}(t_{j-1})$ is the probability of survival (failure-free) prior to $t_j$, given that the system has not experienced any failure mode up to $t_{j-1}$. The term $\hat{h}_i(t_j)$ is the estimated sub-hazard for the failure mode $i$ at $t_j$.

It follows from equation (5.5) and equation (5.8) that the estimated CIF for the system that is experiencing the failure mode $i$ at the time $t$, is given by:

$$\hat{F}_i(t) = \sum_{\forall\, j, t_j \leq t} \hat{h}_i(t_j)\hat{S}(t_{j-1}) \tag{5.9}$$

It is noticed from equation (5.9) that the probability of failing (the incidence) due to the failure mode $i$ at time $t_j$, is simply equal to the probability of surviving at the previous time period $t_{j-1}$ multiplied by $\hat{h}_i(t_j)$. Hence, it is called cumulative incidence function. Since it has a cumulative nature, it is a monotonically increasing function. The CIF estimates the marginal probability for the failure mode $i$ as the cumulative sum up to the time $t_j$ of these incidence values, over all failure times of that failure mode. In other words, the CIF is the sum of the probabilities of observing failure mode $i$ until time $t_j$, while the system is still at risk: that is to say, the system did not experience any failure prior to $t_j$.

The sub-hazard $\hat{h}_i(t_j)$ for the failure mode of type $i$ in equation (5.9), is estimated as given in [13] for the proportion of systems failing due to that failure mode, as: $\hat{h}_i(t_j) = \frac{d_{ij}}{n_j}$, where $d_{ij}$ is the number of systems that have experienced this type of failure at $t_j$, and $n_j$ is the number of systems at risk (those that have not experienced any failure mode before $t_j$). It implies:

$$\hat{F}_i(t) = \sum_{\forall\, j, t_j \le t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1}) \tag{5.10}$$

This means that the CIF estimator for the failure mode $i$ depends not only on the number of systems that have experienced this type of failure ($d_{ij}$), but also on the number of systems that have not experienced any other failure mode ($n_j$). Therefore, equation (5.10) represents the probability that a system will experience a failure mode $i$ by time $t$, in the presence of the other competing failure modes.

The CIF requires that the overall hazard is the sum of the individual hazards for all failure modes [8]. Consequently, the overall cumulative function is equal to the sum of the CIFs for all failure modes, as follows:

$$\hat{F}(t) = \sum_{i=1}^{K} \hat{F}_i(t) \tag{5.11}$$

### 5.4.2 Limitation of the KM estimator in the presence of competing failure modes

The KM estimation method for analyzing the data collected from a set of systems subjected to multiple failure modes, is based on the analysis of each failure mode separately, while the other competing modes are treated as censored. The estimated KM survival curve for each failure mode has questionable interpretation when the other modes are treated as censored. This is because the complement of the KM estimate denoted by $\hat{F}(t) = 1 - \hat{S}(t)$, at any time $t$, is larger than the estimate of each of the CIF curves. The proof is presented in [13]. This means that the estimated survival probability using the KM in the presence of competing failure modes is not informative.

The analysis of this situation can be interpreted as follows: the survival probability using the KM estimation is mainly based on the independent censoring, which requires that those systems which are censored due to other failure modes, at a certain time, are as likely to have a failure later as those in the risk set.

Unlike the estimated KM survival curve, the assumption of independence between the competing failure modes, is not required for the estimation of the CIF. Therefore, the estimated CIF is informative and has clear interpretation for the incidence probability of a given failure mode in the presence of the other competing modes.

## 5.5 Multi-Class Logical Analysis Of Data

LAD is a supervised machine learning classification technique that relies on extracting the knowledge from the training dataset in the form of interpretable patterns. Those interpretable patterns represent the interactions between the covariates for each class of observations in the training dataset. The extracted patterns are then used to formulate a decision model that classifies the unseen data to one of the classes.

### 5.5.1 Stages of the LAD approach and characteristics of the patterns

The LAD approach is composed of three stages: data binarization, pattern generation, and theory formation [14]. The data binarization stage involves the transformation of each numerical observation vector into a vector of binary attributes. The pattern generation stage is essential in identifying the positive and negative patterns from the binarized dataset. It is the cornerstone in LAD approach. The two-class LAD which generates an entire set of patterns is discussed in details in [15]. In the theory formation stage of the two-class LAD, the generated patterns are used to create a decision model called the discriminant function that generates scores for the tested observations, in order to classify them into the positive and negative classes. The accuracy of the LAD decision model depends on the type of the generated patterns [16].

A pattern is said to cover a certain observation if it is true for that particular observation [14]. The set of observations covered by the pattern $p$ is denoted by $Cov(p)$. A pattern can cover some observations from one class, but cannot cover any observations from the opposite class. This is called *pure* pattern [17]. This definition is relaxed to allow the pattern to cover a large proportion of observations in one class, and a much smaller proportion of the observations in the opposite class (this is called *non-pure* pattern) [18]. For more details about LAD, the interested readers may be referred to the comprehensive material that is found in [14].

### 5.5.2 Multi-class LAD decision model

The multi-class classification problem (*polychotomizer)* can be obtained by combining the conventional two-class classifiers in several methods [19]. A multi-class LAD method, that involves modifying the architecture of the two-class LAD, is proposed in [20]. In that work, a unified multi-class LAD classifier is obtained by combining a set two-class LAD classifiers. The method aims at modifying the architecture of LAD as a dichotomy to a multi-class decision model. The proposed method has the advantage that it generates a less complex decision model, in a better execution time. Given a training dataset that consists of different classes of observations, the knowledge is discovered in the form of sets of generated patterns that are representing those classes.

The discriminant function used in the multi-class LAD is significantly different than that of the two-class LAD. Given a new testing observation that is not found in the training data, the discriminant function in multi-class LAD generates a score for each class and therefore the testing observation belongs to the class with the highest score. The score is calculated for a certain class $i$, by using all the pattern sets $P_i$ ($i \in \{1,2,..C\}$). Each set $P_i$ separates class $i$ from all the remaining $(C-1)$ classes. The score for the testing observation $O$ is calculated as given in [21] as follows:

$$\hat{\Delta}(O) = \underset{i=1,\ldots C}{arg\ max} \sum_{p_t^i \in P_i} p_t^i(O) W_t^i \tag{5.12}$$

where $p_t^i(O) = 1$ if the pattern $p_t^i$ covers the observation $O,$ and zero otherwise. The values $W_t^i$ associated with each pattern $p_t^i$ in the set $P_i$, act as normalized weights. The output of equation (5.12) is the class with the highest score. More details are found in [21] to show how to generate a set of multi-class patterns that can be used to create the decision model in the multi-class LAD approach.

## 5.6 The Proposed Multiple Failure Modes Prognostic Methodology

The proposed prognostic methodology based on LAD and CIF is presented in this section. The objective is to predict the RUL of a set of systems working under different operating conditions, in the presence of different competing failure modes. The main concept is to merge both LAD as a multi-class diagnostic technique, and a set of CIFs. As stated in Section 5.4, each CIF represents the marginal probability function for one failure mode, in the presence of the other modes as competing ones. The CIF curve represents an estimate that reflects the effect of the age on the health state of the system. The generated patterns by using the multi-class LAD reflect the effects of the covariates on the health state of the monitored system, and on the interactions among the failure modes.

The proposed methodology consists of two phases; training and updating. Each phase consists of two steps. In the training phase, the failure times for the distinct failure modes are used to estimate a set of CIFs, one CIF for each failure mode. The multi-class LAD is used as diagnostic technique to extract the hidden knowledge from the covariates of the condition monitoring data, in the form of the generated patterns for each failure mode. The set of generated patterns are then refined into a set of selected patterns, by using a pattern selection procedure. In the updating phase, the CIF curves for the failure modes and the generated patterns, are used to update the survival curves of a similar type of monitored systems, in order to obtain certain prognostic indices. After collecting the most recent observation from the monitored system, the covering patterns are identified and the discriminant function score is obtained by using the multi-class LAD diagnostic model. The obtained score is then used to update the survival curve of the monitored system and to predict its RUL. Figure 5-2 shows the phases and the corresponding steps of the methodology. In what follows, the details of each step, are presented.

**Step 1. Estimation of the CIF curves**

A CIF curve is estimated for the failure mode $i \in \{1,2,..C\}$, by using the historical lifetime data collected from the systems that failed due to that failure mode. As depicted in Figure 5-2, the survival function is estimated first for all the systems, by considering all the lifetime data using the KM estimator in equation (5.3), then a CIF curve is estimated for each failure mode by using equation (5.10). Each CIF curve extracts the characteristic information of the corresponding failure mode. So far, each of the estimated CIF curves does not reflect the effect of the operating conditions on the system's health state.

**Step 2. Pattern generation and selection for each failure mode using multi-class LAD**

The multi-class LAD is used to generate sets of patterns from the historical condition monitoring data. These historical data contain a number of observations collected from a set of systems. Each system starts running from its new state (good condition) until the occurrence of complete failure. Each system failed due to one of many failure modes. A set of patterns $P_i$ for the failure mode $i \in \{1,2,..C\}$, are generated from all the observations collected from the systems that failed due to that failure mode. Accordingly, each patterns' set characterizes the corresponding failure mode. These sets of patterns reflect the effect of changing the values of the covariates of different failure modes on the failure time of the monitored system.

The generated patterns represent the interaction between the covariates of such historical data. Each generated pattern is allowed to cover a large proportion of observations of the corresponding class (observations from a given failure mode), and small proportion of observations from the other classes (the competing failure modes). A pattern selection procedure is carried out to select the significant patterns in each set and to remove the redundant ones. In this prognostic methodology, the pattern generation and selection procedures are carried out by using the software *cbmLAD* [22].

**Step 3. Updating the survival curve of the monitored system**

In the updating phase of this methodology, the LAD diagnostic model is used to detect the patterns that are hidden in each new observation, so that the prognostic estimates can be updated appropriately. Based on a set of observations collected recently from a set of similar type of systems, the survival curve for each system is updated according to the patterns covering those observations. The covering patterns reflect the effect of the covariates and the competing failure modes on the system's survival curve. If the recent observation is covered by a set of patterns that belong to a certain failure mode, this mode consequently affects the survival curve of the monitored system. Updating the system's survival curve is simply explained in the following.

After collecting the current observation from the monitored system, the multi-class LAD detects the covering patterns for that observation. The covering patterns are used to calculate the score using the discriminant function in equation (5.12). Based on the obtained highest score, the survival curve of the recent observation $O$ that belongs to class $i$ at time $t_k$, is updated using the following updating formula:

$$S_{O(t_k)}(t) = \begin{cases} q_i - \hat{F}_i(t) & \forall \;\; k = 1 \\ [q_i - \hat{F}_i(t) + S_f(t)\,]/2 & \forall \;\; k \geq 2 \end{cases} \tag{5.14}$$

where $S_f(t)$ is the former updated survival curve based on the previous observation $O(t_{k-1})$. The updating formula can be simply clarified as follows: given the first observation collected from the system, its survival curve is updated by considering the sub-survival curve $\hat{S}_i(t) \; = q_i - \hat{F}_i(t)$ for the failure mode $i$ which is identified based on the score calculated by the multi-class LAD diagnostic model. The subsequent observations are then employed to update the survival curve, by adding the sub-survival curve of the identified failure mode to the former updated curve $S_f(t)$, and then calculate the average. Considering the former survival curve makes the prognostic model be able to reflect the history of the monitored system.

**Step 4. RUL calculation based on the updated curve**

In this prognostic methodology, the updated survival curve in step 3 is used to reflect the effect of the operating conditions and the competition between the failure modes on the RUL of the system. The time to failure is represented previously as a continuous random variable and denoted by $T$. The RUL of the system at the monitoring (inspection) time $t_k$ is also random variable denoted by $T - t_k$. The expected value of the RUL called the *mean remaining useful life* (*MRUL*), is calculated as presented in [23], by considering only one failure mode, as follows:

$$MRUL(t_k) = E(T - t_k | T > t_k) = \frac{\int_{t_k}^{\infty} (\tau - t_k) f(\tau) d\tau}{S(t_k)} \tag{5.15}$$

The MRUL can be calculated experimentally, by using the estimated KM survival function $\hat{S}(t)$ in equation (5.3). Accordingly, equation (5.15) is represented in the discrete form as:

$$MRUL(t_k) = \frac{\sum_{t_r = t_k}^{\infty} t_r \left[\hat{F}(t_{r-1}) - \hat{F}(t_r)\right]}{\hat{S}(t_{k-1})} - t_k \tag{5.16}$$

where $\hat{F}(t) = 1 - \hat{S}(t)$.

By considering the sub-survival of each failure mode, the remaining useful life for each sub-distribution (or the MRUL for the cause-specific failure mode $i$) is calculated, by considering the sub-density function $f_i(t)$ and the sub-survival $S_i(t)$, as follows:

$$MRUL_i(t_k) = E(T - t_k, i | T > t_k, i) = \frac{\int_{t_k}^{\infty} (\tau - t_k) f_i(\tau) d\tau}{S_i(t_k)} \tag{5.17}$$

Equation (5.17) is represented experimentally as:

$$\begin{aligned} MRUL_i(t_k) &= \frac{\sum_{t_r = t_k}^{\infty} t_r \left[\hat{F}_i(t_r) - \hat{F}_i(t_{r-1})\right]}{\hat{S}_i(t_{k-1})} - t_k \\ &= \frac{\sum_{t_r = t_k}^{\infty} t_r \left[\hat{S}_i(t_{r-1}) - \hat{S}_i(t_r)\right]}{\hat{S}_i(t_{k-1})} - t_k \end{aligned} \tag{5.18}$$

The estimated sub-survival function $\hat{S}_i(t)$ in equation (5.18) is derived from the estimated CIF for the failure mode $i$, by using equation (5.6).

Based on the collected updating observation and the resulting updated survival curve, the $MRUL$ of the monitored system is calculated as:

$$\widehat{MRUL}_{O(t_k)} = \frac{\sum_{t_r=t_k}^{\infty} t_r\left[s_{O(t_k)}(t_{r-1}) - s_{O(t_k)}(t_r)\right]}{s_{O(t_k)}(t_{k-1})} - t_k \tag{5.19}$$

In the next section, a case study is employed to test the effectiveness of the proposed methodology.

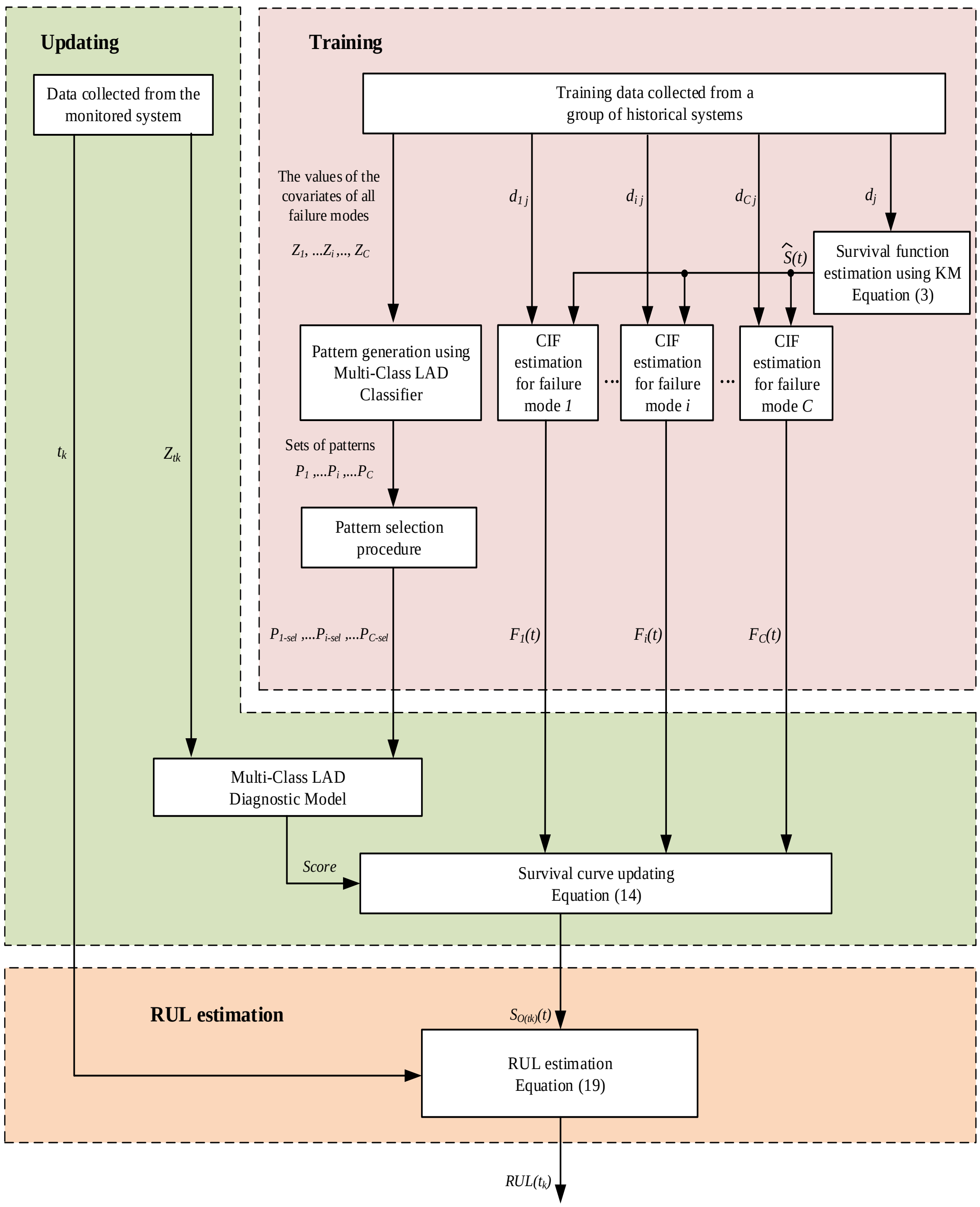


Figure 5-2: Phases and steps of the proposed multiple failure modes prognostic methodology

## 5.7 Case Study: Rotating Machinery Application

### 5.7.1 Multiple failure modes prognostics in rotating machinery

The rotating machinery is widely used in a multitude of industrial systems, including naval and automotive industries, aircraft engines in aeronautics, power plants' turbines, and train transmission systems [24, 25]. Components such as bearings are at the heart of rotating machinery and are the most likely to have failures. They are commonly used in rotating machines to support rotating shafts and to transmit torque; they are the major causes of the rotating machinery catastrophic failures that result in costly downtime. Incipient faults that occur in the bearings are usually caused by localized defects in the races, or the rolling element. The failure of the bearings can result in the deterioration of the health state of the rotating machine [26].

The problem of multiple failure modes prognostics in the rotating machinery can be summarized in the following. At the system level, different components can have different failures such as bearing defect, cracked or broken rotor bars, mechanical seal wear, and others. At the component level, there may be different types of failure such as the bearing inner race defect, outer race defect, a crack in the cage of rolling element.

In fact, the rotating machinery prognostics pose important challenges in the reliability and maintenance fields, in particular when there are competing failure modes. It is important to develop a multiple fault diagnostic/prognostic scheme to identify the different faulty patterns of the rotating machinery and to predict the failure time as well. The objective is to decrease the downtime of the machines in production and to increase their reliability against possible competing failures. It is therefore necessary to accurately and automatically diagnose the existence of incipient faults in the bearings, and to predict the RUL of such significant components.

### 5.7.2 Prüftechnik Canada vibration data

One of the most common tools for fault diagnostic and prognostic in rotating machinery is the vibration analysis [1, 27]. The proposed LAD prognostic methodology explained in the previous section is tested and validated on a database of vibration signals obtained from bearing test rigs in *Prüftechnik Canada* [28]. The bearings started running from the brand new state (good condition). Three types of faults have been initialized in bearings by seeding three types of defects; inner race defect, outer race defect, and rolling element defect. The test was carried out until the complete failure of each bearing. The vibration signals were acquired continuously from each bearing by the accelerometers at a sampling frequency of 65384 Hz, for a time record length of 62 milliseconds, between different inspection intervals of 40-90 minutes. Each time record of the vibration signals containing 4096 sampling points, taken per accelerometer snapshot. The collected vibration signals carry the information about the status of the bearings from the good condition state throughout their deterioration states until the occurrence of complete failure.

### 5.7.3 Feature Extraction

It is extremely difficult for the vibration analyst to identify the faults in a monitored bearing through the time domain data. This is because the vibration signals are non-stationary time function that changes in both amplitude and distribution, and does not give a clear information about the type of faults being identified [1].

Moreover, a vibration signal generated by a certain bearing fault has relatively low amplitude, and it is often overwhelmed by noises with higher amplitudes and other sources of vibrations generated from other components in the rotating machine. Therefore, in order to exploit the information found in the raw signal efficiently, it is necessary to extract the faulty features that identify the different failure modes. Thus, the raw signals are preprocessed and some statistical metrics, which represent the features are calculated.

In this paper, five statistical features are calculated from the time domain data and are considered as covariates. The following features are calculated for the signal $x(n)$: the kurtosis, root mean square (RMS), crest factor, skewness, and peak-to-peak value. These feature are listed in Table 1.

Table 5.1: Time domain-based features

| Feature | Mathematical Formula |
|---|---|
| Kurtosis | $kurtosis(x(n)) = \frac{\sum_{n=1}^{N}[x(n)-\mu]}{(N-1)\sigma^4}$ <br> where $\mu$ and $\sigma$ are the mean and standard deviation of the signal $x(n)$, respectively and $N$ is the length of the time record of the signal. |
| Root Mean Square (RMS) | $RMS(x(n)) = \sqrt{\frac{\sum_{n=1}^{N}[x(n)]^2}{N}}$ |
| Crest Factor | $Crest\ Factor(x(n)) = \frac{peak(x(n))}{RMS(x(n))}$ |
| Skewness | $Skewness(x(n)) = \frac{\sum_{n=1}^{N}[x(n)-\mu]}{(N-1)\sigma^3}$ |
| Peak-to-Peak | $Peak - Peak(x(n)) = Max(x(n)) - Min(x(n))$ |

It is important to reveal the temporal structure of the vibration signal; how the signal's frequency contents vary with time. This can help to clarify if two or more frequency components are found throughout or in some intervals of the signal. Thus, it is necessary to represent the raw signals in the time-frequency domain, in order to handle the non-stationary nature of such signals.

In this paper, the wavelet transform (WT) is suggested as a multiscale time-frequency analysis. It is used to represent the non-stationary signals through dilation and translation processes. The WT uses a series of oscillating functions with different frequencies as window functions to scale and translate the time domain signal. These functions are derived from a short wave called the mother wavelet function. The continuous wavelet transform (CWT) equation presented in [29], is expressed as:

$$W_{\psi}(a,b) = \int_{-\infty}^{\infty} x(t)\psi^{*}(\frac{t-b}{a})dt \tag{5.20}$$

where $x(t)$ the continuous time domain signal, $a$ is the scale parameter and $b$ is the translation parameter of the mother wavelet function $\psi^{*}(t)$ along the time axis. There are many types of mother wavelet functions and each one has its own applications and merits. The most common mother wavelets are the Morlet, Daubechies, Haar, and Mexican Hat wavelet [30]. The mother wavelet is given by:

$$\psi_{a,b}^{*}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{5.21}$$

The WT is basically expressed as the dilation and the compression of the translated wavelet. The main idea behind this dilation or compression is that if the wavelet is dilated it represents the low frequencies in the time domain signal, and if it is compressed it has a high rate of change and as a result it represents the high frequencies in the signal [29].

Since the vibration signal has a discrete nature rather than continuous, therefore there is a need to perform the discrete wavelet transform (DWT). The DWT for the discrete time signal $x(n)$ is represented in the following equation [30]:

$$DWT_{\psi}(a,b) = 2^{-\frac{a}{2}} \sum_{n} x(n)\psi(2^{-a}n-b) \tag{5.22}$$

The objective of the DWT analysis is to decompose the signal into a set of frequency bands, by using a set of low-pass and high-pass filters, along with a decimation procedure (it is the process of reducing the sampling rate of a signal) [30, 31]. As a result of that decomposition, the time domain signal is converted into a set of wavelet coefficients. These coefficients represent the approximation and details in the time domain signal [29].

In this paper, the Wavelet Toolbox in MATLAB [31] is used to carry out the DWT decomposition for the vibration signals in this case study. Ten time-frequency domain scales are analysed using the Daubechies wavelet (db10). The input to the Wavelet toolbox is the vibration signal collected from the bearing, and the output is a scalogram consisting of ten scales (levels). The scalogram provides a clear image visualization for the energy levels in the signal. Consequently, it is easier to compare and identify abnormalities or anomalies in the faulty signals that are coming from different failure modes. Each of the scales in the scalogram represents the detailed information about the defect frequencies (the different abnormalities in each failure mode).

So far, to able to automatically diagnose the existence of certain failure mode in the monitored bearing, suitable features can be extracted either using image-processing techniques or by calculating the energy values in each scale in the wavelet scalograms, by using the wavelet coefficients directly [29]. In this paper, the wavelet analysis is carried out in MATLAB programming environment to calculate the relative energy in each scale in the wavelet scalogram, in order to use it as a fault-feature to train the multi-class LAD classification algorithm. For each raw signal, a fault-feature vector is extracted to represent the percentages of energy in all levels of the wavelet scalogram.

In this paper, the five time-domain features in addition to the ten wavelet scales-based features (WS-1 to WS-10), are respectively extracted from each collected vibration signal, to train the multi-class LAD classifier. A sample of the processed observations in the training data for the three failure modes, is displayed in Table 5.2. Each observation is represented as a numerical vector that contains the inspection time and the covariates (the fifteen extracted features). The objective is to generate a set of patterns that are recognising the various failure modes of the bearings.

Table 5.2: A sample of the processed observations collected from three different bearings having three different failure modes

| Failure mode | Time (Hours) | Covariates | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Time domain Features | | | | | Time-Frequency domain (Wavelet-based) Features | | | | | | | | | |
| | | Kurtosis | RMS | Crest Factor | Skewness | Peak-Peak | WS-1 | WS-2 | WS-3 | WS-4 | WS-5 | WS-6 | WS-7 | WS-8 | WS-9 | WS-10 |
| 1 | 57 | 21.298 | 21.197 | 9.578 | 0.969 | 361.208 | 6.217 | 19.524 | 43.186 | 18.242 | 11.524 | 0.794 | 0.254 | 0.192 | 0.058 | 0.003 |
| 1 | 123 | 16.973 | 33.531 | 8.426 | 0.950 | 478.557 | 8.089 | 21.527 | 44.992 | 16.625 | 7.645 | 0.764 | 0.303 | 0.037 | 0.015 | 0.001 |
| 1 | 187 | 18.215 | 40.035 | 8.739 | 0.985 | 595.023 | 9.028 | 24.470 | 44.692 | 15.152 | 5.691 | 0.680 | 0.172 | 0.075 | 0.038 | 0.002 |
| 1 | 202 | 11.205 | 83.453 | 7.2587 | 1.032 | 954.606 | 10.168 | 27.646 | 45.457 | 11.926 | 4.045 | 0.527 | 0.125 | 0.059 | 0.041 | 0.0018 |
| 2 | 65 | 40.892 | 9.840 | 10.916 | 1.0815 | 192.172 | 5.183 | 15.368 | 36.629 | 22.880 | 18.169 | 1.014 | 0.523 | 0.195 | 0.019 | 0.0052 |
| 2 | 134 | 19.075 | 22.019 | 8.024 | 0.715 | 342.875 | 7.886 | 25.632 | 46.149 | 15.544 | 3.949 | 0.557 | 0.228 | 0.028 | 0.023 | 0.0039 |
| 2 | 179 | 16.325 | 27.816 | 7.663 | 0.450 | 419.024 | 9.294 | 24.692 | 36.369 | 20.292 | 8.009 | 0.837 | 0.345 | 0.124 | 0.022 | 0.011 |
| 2 | 216 | 13.683 | 31.951 | 6.949 | 0.304 | 434.443 | 9.086 | 25.241 | 41.871 | 17.971 | 5.019 | 0.598 | 0.189 | 0.017 | 0.0112 | 0.002 |
| 2 | 258 | 11.526 | 40.735 | 6.094 | 0.076 | 490.587 | 9.258 | 25.622 | 43.163 | 12.723 | 8.254 | 0.568 | 0.278 | 0.087 | 0.022 | 0.013 |
| 3 | 78 | 68.591 | 13.093 | 14.191 | 1.913 | 344.394 | 7.166 | 20.108 | 27.509 | 29.638 | 13.433 | 1.672 | 0.397 | 0.056 | 0.008 | 0.0009 |
| 3 | 163 | 25.953 | 43.778 | 10.161 | 0.827 | 802.297 | 8.439 | 25.662 | 46.655 | 13.669 | 4.819 | 0.556 | 0.147 | 0.022 | 0.020 | 0.0094 |
| 3 | 341 | 14.298 | 107.92 | 7.998 | 0.793 | 1515.969 | 12.332 | 32.007 | 43.582 | 9.177 | 2.448 | 0.361 | 0.051 | 0.027 | 0.012 | 0.0055 |
| 3 | 347 | 18.077 | 101.894 | 8.820 | 0.949 | 1595.323 | 13.315 | 31.102 | 43.349 | 9.128 | 2.475 | 0.505 | 0.089 | 0.021 | 0.011 | 0.0032 |

## 5.7.4 Feature Selection

From the CBM fault diagnosis point of view, some features are irrelevant or non-significant [2]. If the whole feature set is employed by the classification technique directly, it may give a lower diagnosis accuracy, and accordingly affects the performance of the prognosis. Thus, to improve the diagnosis accuracy in the proposed prognostic methodology, a set of significant features which obviously discriminate between the data in the different failure modes need to be selected from the original feature set.

Some of the feature selection techniques are applied according to the experience accumulated by different researchers [32, 33]. A feature selection technique called compensation distance evaluation technique (CDET) proposed in [25] is adopted in this paper due to its tractability and reliability. A great advantage of the CDET over the other feature selection techniques, is that it selects the significant features automatically without relying on human experience. Another reason for choosing CDET in this paper is that it does not lose the meaning of the features in the reduced set, and therefore it keeps the interpretability of the LAD decision model which is very important for the CBM decision maker.

It is two-stage feature weighting algorithm that generates a normalized weight (ratio) for each feature. The CDET algorithm takes into account the different importance degrees of all features. A weight for each feature is computed and assigned according to its sensitivity and importance in classification [25]. These weights not only highlight the importance of sensitive features but also reflect the interference of insensitive features.

For the fifteen extracted features in Table 5.2, a subset of features are selected, based on the selection criteria given in [25]. All features that have normalized ratios exceeding a certain threshold value, are selected. In this paper, a threshold value of 0.2 is assigned and accordingly ten features are selected as shown in Figure 5-3. As we will see later, the selected features can give better prognostic results.

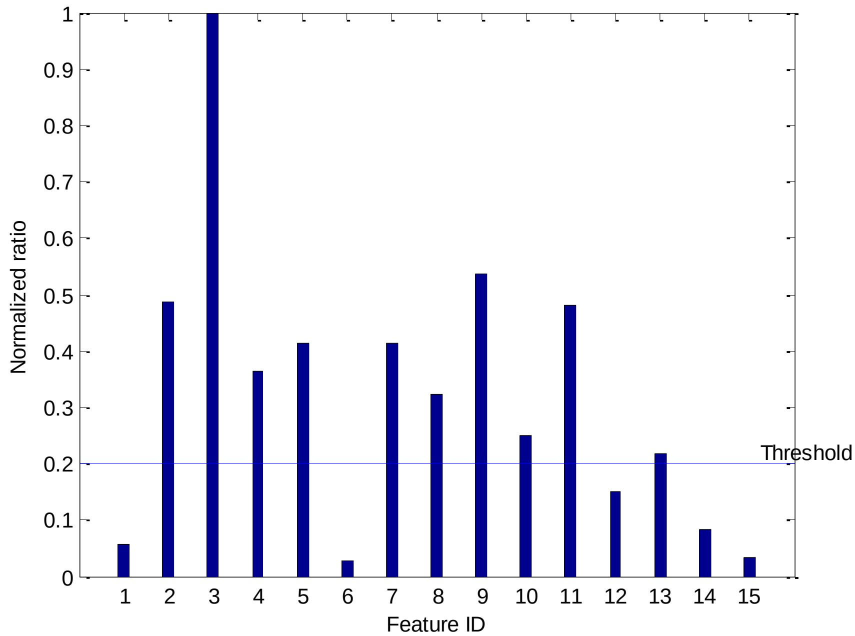


Figure 5-3: Feature selection using CDET

### 5.7.5 CIF Estimation for Each Failure Mode

Based on the lifetime data collected from each failure mode, a CIF curve is estimated for each mode by using equation (5.10). Three CIF curves are estimated; one curve for each failure mode, as depicted in Figure 5-4. These curves represent the marginal probabilities that are reflecting the operational time on the age of the monitored bearing. However, they do not reflect the effects of the extracted features on the health state of the bearing.

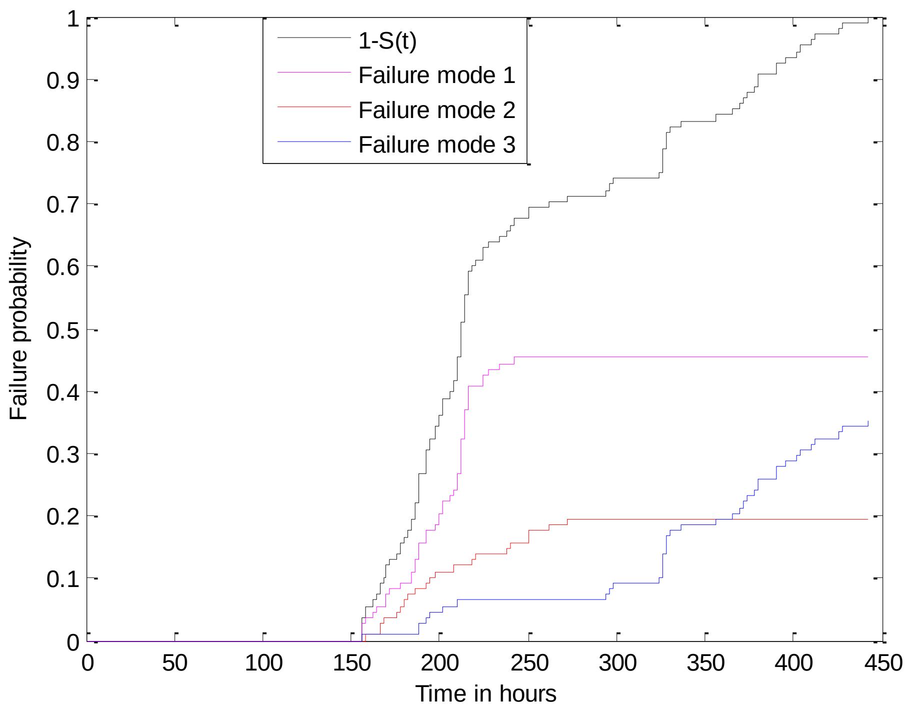


Figure 5-4: CIF for each failure mode and the curve $\hat{F}(t) = 1 - \hat{S}(t)$

### 5.7.6 Pattern Generation Using Multi-Class LAD

The multi-class LAD is used to generate the patterns for each failure mode, by exploiting the extracted features in training dataset. A set of the significant patterns are then selected. The pattern generation procedure is carried out by using the software *cbmLAD* [22]. The generated patterns have an important advantage which is the ease of interpretation. Each pattern is meaningful and has a physical interpretation of the characteristics of the failure mode of monitored bearing. The interpretability of the patterns means that it is easy to evaluate the failure modes and directly relates them to the covariates.

Based on the fifteen covariates in the training dataset, twenty-two patterns are generated. Nine generated patterns for the first failure mode, seven for the second, and six for the third. The generated patterns for each failure mode and their interpretation are listed in Table 5.3. In order to study the effect of feature selection on the performance of the prognostic methodology, another set of patterns are generated from the set of selected features, in the similar way. The number of generated patterns in this case is eighteen patterns.

Table 5.3: Interpretations of the generated patterns (1: Inner race defect, 2: Outer race defect, and 3: Rolling element defect)

| Class | Pattern $p_t^i$ | Time domain Features | | | | | Time-Frequency domain (Wavelet-based) Features | | | | | | | | | | Weight $W_t^i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Kurtosis | RMS | Crest Factor | Skewness | Peak-Peak | WS-1 | WS-2 | WS-3 | WS-4 | WS-5 | WS-6 | WS-7 | WS-8 | WS-9 | WS-10 | |
| 1 | $p_1^1$ | < 13.247 | > 5.1661 | | | | | | > 36.7 | | | | | | | | 0.1165 |
| 1 | $p_2^1$ | > 2.718 AND < 25.891 | | > 3.837 | | | | | | < 23.60 | < 7.74 | | | < 0.67 | | | 0.1098 |
| 1 | $p_3^1$ | < 25.891 | | < 10.017 | > 0.803 | < 1356.46 | | > 20.89 | | | < 8.69 | > 0.292 | | | | | 0.1128 |
| 1 | $p_4^1$ | < 12.756 | | | > 0.0017 | | | > 18.25 | | < 32.65 | < 14.79 | | | < 0.64 | | | 0.1143 |
| 1 | $p_5^1$ | < 12.756 | | | > 3.629 | | | | | < 32.65 | | | | | | | 0.1024 |
| 1 | $p_6^1$ | < 25.891 | | < 10.017 | > 0.746 | > 361.014 AND < 1356.46 | > 6.155 | | > 36.7 | < 23.60 | | > 0.292 | | | | | 0.1128 |
| 1 | $p_7^1$ | < 25.891 | | | > 0.756 | < 1356.46 | | > 21.06 | | | > 3.27 AND < 8.74 | | | | | | 0.1098 |
| 1 | $p_8^1$ | > 2.938 AND < 13.247 | | | | | | | | < 31.83 | < 14.79 | | | < 0.64 | | | 0.1113 |
| 1 | $p_9^1$ | < 13.188 | | | > 0.022 | | > 4.665 | | | | | | | | | | 0.1098 |
| 2 | $p_1^2$ | | | | | < 435.729 | | | | | | < 1.782 | | | | | 0.1603 |
| 2 | $p_2^2$ | | | | | < 511.267 | | < 25.94 | | | | | > 0.18 | | | < 0.0234 | 0.1526 |
| 2 | $p_3^2$ | | | | | < 472.408 | | | < 40.57 | > 19.29 | > 7.45 | | | | > 0.003 | < 0.1586 | 0.1371 |
| 2 | $p_4^2$ | > 11.511 | | | | < 511.267 | | | | | | | | | | | 0.1806 |
| 2 | $p_5^2$ | | | | | | | < 20.31 | | | > 7.45 | | | | < 0.427 | | 0.1072 |
| 2 | $p_6^2$ | | | | > 0.0064 | < 511.267 | | | | > 17.79 | | | | | > 0.003 AND < 0.295 | | 0.1371 |
| 2 | $p_7^2$ | | | | | < 472.408 | | | < 40.57 | > 20.33 | | | | | > 0.003 | < 0.1586 | 0.1246 |
| 3 | $p_1^3$ | | | > 3.174 AND < 3.620 | | | | | | | | | | | | | 0.1593 |
| 3 | $p_2^3$ | | | > 3.996 | | | | | | | < 10.49 | > 0.901 | | | | | 0.0938 |
| 3 | $p_3^3$ | | < 111.89 | > 3.174 | | | | | > 38.46 AND < 43.69 | | | < 1.258 | | | | | 0.2304 |
| 3 | $p_4^3$ | | | > 7.996 | | > 472.408 | | | < 49.27 | | | | | | | | 0.2475 |
| 3 | $p_5^3$ | | | > 9.392 | | | | | | | < 13.50 | | | | | | 0.1806 |
| 3 | $p_6^3$ | | | | < 0.0146 | | | | | | | | | | | | 0.0882 |

### 5.7.7 Validation of the Proposed Methodology

The proposed methodology is validated and its performance is compared to the performance of two of the most common machine learning techniques; the artificial neural networks (ANNs) and support vector regression (SVR). ANNs have desirable characteristics such as generalization ability, learning ability, and adaptability [34]. Such networks are used commonly in the CBM fault prognostics to represent the nonlinear relationship between the covariates and the RUL of the monitored equipment [2, 3, 35].

**ANN and SVR prediction models**

In this paper, the ANN prediction model is trained using the processed data that are collected from the bearings. The network topology is shown in Figure 5. It consists of three layers; the input layer, hidden layer, and output layer. The number of input nodes is fifteen that represent the extracted features (the five time domain and the ten wavelet-based features). The optimal number of neurons in the hidden layer is not easy to determine from the first time and needs a number of experimentations to be tuned. In this paper, the optimal number of neurons in the hidden layer is set to nine. The network is trained by using the observations in the training dataset, by providing the actual RUL to the output neuron as a desired output, and the corresponding features to the input layer. The actual RUL is the difference between the failure time of the bearing and its current age. Two ANN models are obtained. The first model is the one trained using all the fifteen extracted features. The second model is trained using the selected ten features according to the CDET algorithm. The weights of each ANN model are adjusted by using gradient-descent method [36], based on the difference between the actual and estimated RUL of the bearings. The trained models extract the relationship between the RUL and the extracted features.

The second prediction model which is compared to the proposed methodology is the SVR. It is based on the concept of support vector machine (SVM), which was presented by Vapnik [37]. The SVR prediction model has two main important advantages over the ANN prediction model. First, it has better generalization ability due to choosing the maximal margin hyperplane, which results in minimizing the risk of overfitting [36]. Second, it can provide more complex nonlinear functions by applying the kernel trick [37]. A kernel mapping is applied to transform the original covariates in the dataset into higher dimensional space, and accordingly the SVR becomes a nonlinear function in the original covariates.

Several kernels can be applied depending on the application at hand. Polynomial kernel, radial basis function (RBF) kernel, and sigmoid kernel are commonly used. In this paper, the RBF kernel is chosen, and it gives the best results among the other kernels. Similar to the training process of the ANN, two different SVR models are trained. One exploits all the extracted feature and the other uses the selected features only. The ANN and SVR prediction models are implemented in the machine learning software *Weka* [36, 38].

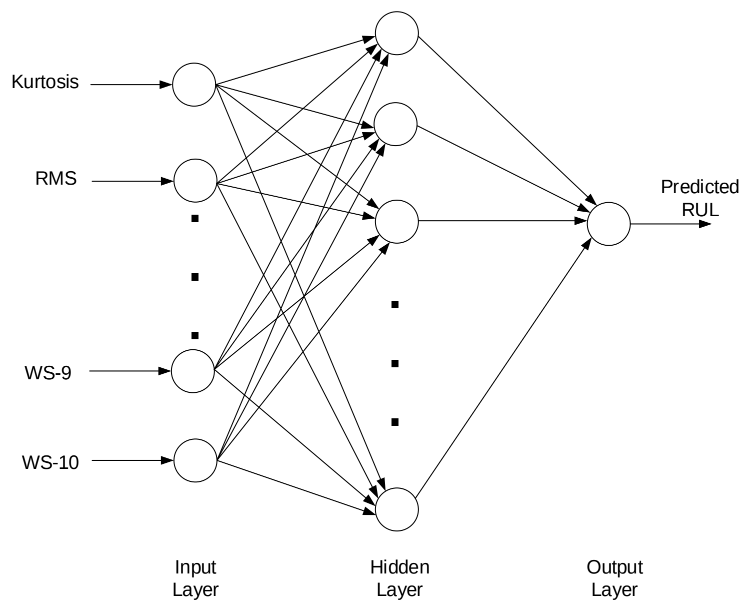


Figure 5-5: The ANN prediction model

**LAD-Kaplan Meier (LAD-KM) Prediction Model**

The proposed LAD methodology is compared with the prognostic methodology presented in [4]. That methodology uses the two-class LAD and the KM estimator to estimate the RUL. The two-class LAD is used in that methodology to reflect the effects of the covariates on the system; health state, and KM is used to estimate the baseline curve. The methodology does not make any assumption about the distribution of the historical lifetime data. This makes it advantageous to be applied in a multitude of CBM prognostic problems. However, it considers the lifetime data that are drawn from a single failure mode. Our hypothesis states that the proposed multiple LAD methodology can predict the RUL more accurately than that methodology, since we consider the competition between the different failure modes.

**Testing vibration data collected from three different failure modes**

The validation is carried out by collecting testing vibration data from another set of bearings that are not found in the training dataset. The testing data are collected from three bearings running to failure with different types of defects; inner race, outer race defect, and rolling element defect. The data are processed and the features are extracted. Given an observation collected from each bearing in the testing dataset, each of the trained models (the proposed LAD, the ANN, the SVR model, and the LAD-KM) is used to predict the RUL instantaneously. The average absolute prediction error is used in this paper to assess the prediction performance of each model. It measures the discrepancy between the actual $RUL$ and the estimated $MRUL$ of each bearing, it is given as:

$$\bar{E} = \frac{1}{N}\sum_{k=1}^{N}\left|RUL(t_k) - \widehat{MRUL}(t_k)\right| \tag{5.23}$$

where $N$ is the actual number of testing observations collected from the monitored bearing.

The results of the calculated average absolute prediction error for each testing bearing, by using each of the trained models, are listed in Table 5.4 and Table 5.5. Table 5.4 presents the results when all the extracted features are inputted to the prediction models, while Table 5.5 presents the case when the set of selected features are used.

From the two tables, the results of the prediction error in the three bearings show that the proposed LAD prognostic methodology is promising and more accurate than the other models, for predicting the RUL in the presence of multiple failure modes. The percentage of the average absolute error to the TTF for each bearing is also presented in the two tables. These percentages indicate that the prediction performance of the proposed LAD methodology does not have a significant change when the life spans of the bearing changes.

It is noticed from Table 5.5 that the feature selection procedure affects the performance of all the prediction models. The application of the selected features gives a better performance for all the models, in particular the ANN model and the SVR model. However, the change in the performance of the proposed LAD methodology is not that significant. This may be attributed to the fact that LAD can deal with a small or large number of features; either correlated or uncorrelated, which is one of the most important advantages of such technique.

Table 5.4: Average prediction error when the prognostic models consider all features in the training dataset

| Test Bearing number | Failure Mode | Time To Failure (TTF) (in hours) | Average prediction error $\bar{E}$ | | | | Percentage of $\bar{E}$ to TTF | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Proposed LAD | ANN | SVR | LAD-KM | Proposed LAD | ANN | SVR | LAD-KM |
| 1 | Inner race defect | 214 | **17.46** | 35.18 | 26.13 | 48.31 | 8.16 % | 16.43 % | 12.21 % | 22.57 % |
| 2 | Outer race defect | 232 | **26.60** | 55.23 | 51.49 | 97.71 | 11.46 % | 23.81 % | 22.19 % | 42.12 % |
| 3 | Rolling element defect | 328 | **31.55** | 47.83 | 44.32 | 86.73 | 9.62 % | 14.58 % | 13.51 % | 26.44 % |

Table 5.5: Average prediction error when the prognostic models consider the selected features only

| Test Bearing number | Failure Mode | Time To Failure (TTF) (in hours) | Average prediction error $\bar{E}$ | | | | Percentage of $\bar{E}$ to TTF | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Proposed LAD | ANN | SVR | LAD-KM | Proposed LAD | ANN | SVR | LAD-KM |
| 1 | Inner race defect | 214 | **15.28** | 25.73 | 19.43 | 44.72 | 7.14 % | 10.02 % | 9.08 % | 20.89 % |
| 2 | Outer race defect | 232 | **22.84** | 47.42 | 43.97 | 94.82 | 9.84 % | 20.44 % | 18.95 % | 40.87 % |
| 3 | Rolling element defect | 328 | **27.55** | 38.79 | 38.42 | 82.30 | 8.39 % | 11.83 % | 11.71 % | 25.09 % |

In addition to the average prediction error, the penalty function defined in [3] is used here as a validation criterion for the proposed prognostic methodology. The prediction accuracy function $L_{Acc}$ measures the difference between the actual $RUL$ and the estimated $RUL$ at the inspection time $t_k$ in each bearing. It is defined as follows:

$$L_{Acc}(t_k) = \begin{cases} \alpha\left[RUL(t_k) - \widehat{MRUL}(t_k)\right] & \text{if} \quad \widehat{MRUL}(t_k) < RUL(t_k) \\ 0 & \text{if} \quad \widehat{MRUL}(t_k) = RUL(t_k) \\ \beta\left[\widehat{MRUL}(t_k) - RUL(t_k)\right] & \text{if} \quad RUL(t_k) < \widehat{MRUL}(t_k) \end{cases} \tag{5.24}$$

where $\alpha$ and $\beta$ are underestimation and overestimation parameters, respectively. As shown in Figure 5-6, the overestimation is penalized more than the underestimation in RUL prediction. Accordingly, the value of $\alpha$ is set to be less than the value of $\beta$. This is more feasible since the overestimation has more severe consequences in the practical situations.

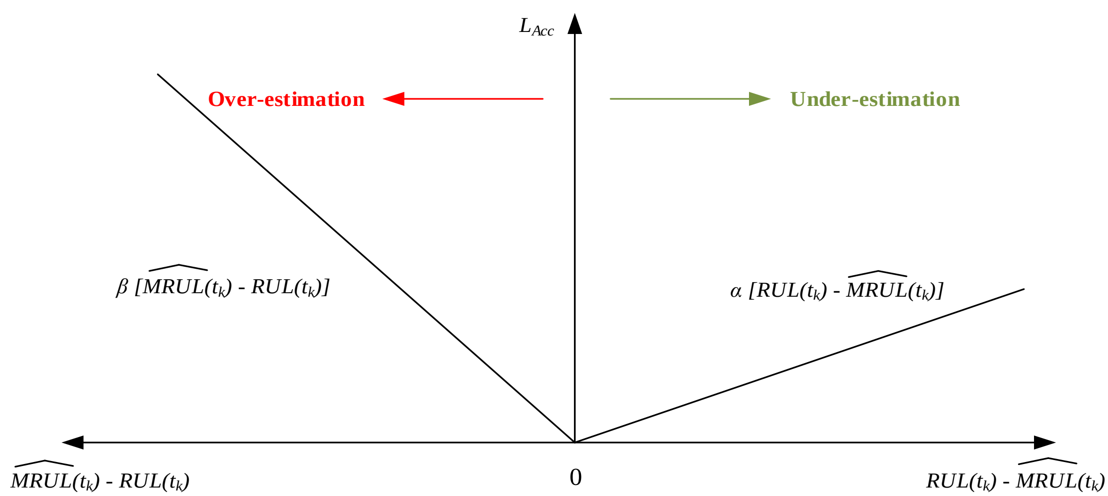


Figure 5-6: The prediction accuracy function

For the bearing $g$, the average penalty incurred is defined as:

$$L_{AP}\left(\text{bearing}_g\right) = \frac{1}{N}\sum_{k=1}^{N}\left[L_{Acc}(t_k\ )\right] \tag{5.25}$$

The total penalty incurred for each prediction model is defined as:

$$L(\text{model}) = \frac{1}{G}\sum_{g=1}^{G}\left[L_{AP}\left(\text{bearing}_g\ \right)\right] \tag{5.26}$$

where $G$ is the number of bearings in the testing dataset. In this paper, the values of $\alpha$, $\beta$ are arbitrarily assigned with 0.1 and 0.3, respectively.

The results of the incurred penalty for each bearing, when all the extracted features are exploited to train all the prediction models, are listed in Table 5.6. Table 5.7 lists the results for the incurred penalty when the set of selected features are used. The obtained results shown in Tables 5.6 and 5.7 indicate that the prediction error of the proposed LAD methodology is penalized less than those of the other prediction models. These results show that the proposed prognostic methodology is promising and can predict the RUL accurately, in the presence of multiple failure modes. The results also show that the proposed methodology has a stable performance, since there is no big change in the incurred penalty due to the variation of the life spans of the bearings.

Table 5.6: The total incurred prediction penalty for each prognostic model when all extracted features are considered

| Prediction Model | Average penalty incurred for the first bearing $L_{AP}(Bearing\ 1)$ | Average penalty incurred for the second bearing $L_{AP}(Bearing\ 2)$ | Average penalty incurred for the third bearing $L_{AP}(Bearing\ 3)$ | Total penalty incurred for each model $L(model)$ |
|---|---|---|---|---|
| Proposed LAD | **2.438** | **3.395** | **4.733** | **3.5220** |
| ANN | 5.095 | 7.934 | 6.315 | 6.4480 |
| SVR | 3.736 | 7.637 | 5.867 | 5.7467 |
| KM-LAD | 6.360 | 14.554 | 13.009 | 11.3077 |

Table 5.7: The total incurred prediction penalty for each prognostic model when the selected features using CDET are considered

| Prediction Model | Average penalty incurred for the first bearing $L_{AP}(Bearing\ 1)$ | Average penalty incurred for the second bearing $L_{AP}(Bearing\ 2)$ | Average penalty incurred for the third bearing $L_{AP}(Bearing\ 3)$ | Total penalty incurred for each model $L(model)$ |
|---|---|---|---|---|
| Proposed LAD | **2.133** | **3.139** | **4.1325** | **3.1348** |
| ANN | 3.645 | 6.771 | 5.2055 | 5.2072 |
| SVR | 2.905 | 6.416 | 5.1730 | 4.8313 |
| KM-LAD | 5.866 | 14.093 | 12.3450 | 10.7680 |

## 5.8 Conclusions

This paper presents a new multiple failure modes prognostic methodology in CBM. The presented methodology combines the multi-class LAD and a set of non-parametric cumulative incidence functions (CIFs). It can be applied to the field of CBM prognostics where a set of systems are working under different operating conditions and are subjected to a multitude of competing failure modes. The multi-class LAD is used to extract the hidden knowledge from the condition monitoring data of different failure modes, by detecting the high degree of interactions among the covariates, without making any priori statistical assumptions.

The proposed methodology consists of two phases; training and updating. In the training phase the knowledge is extracted from the condition monitoring data in the form of relevant patterns and a set of CIF curves that reflect the temporal characteristic information, in the presence of the other competing failure modes. In the updating phase, after collecting a new observation, the diagnostic information obtained by using the multi-class LAD are used to update survival curve of the monitored system and to estimate its RUL as well.

The RUL prediction results show that the prognostic methodology is able to reliably recognize the different failure modes of rotating machinery in the industry. A comparative study is performed between the proposed prognostic methodology and two common machine learning techniques, in order to measure its performance. It is also compared to a non-parametric prognostic methodology in the field of CBM, in order to study the effect of the competing failure modes. The results of comparing the performance of the proposed methodology to all prognostic models show that it is promising and can accurately predict the RUL of the machine in the presence of the multiple failure modes. The results also support our hypothesis that the prediction error of the proposed methodology is penalized less than those of the other models.

The variation in the number of features used for the classification give different prognostic performance. An efficient feature selection algorithm called compact distance evaluation algorithm (CDET) is used to avoid the curse of dimensionality by discarding irrelevant or non-significant features. This algorithm is compatible with the LAD approach. The compatibility means keeping the interpretability of the LAD which is very important for the maintenance decision maker. In this paper, we studied the effect of linking DET to multi-class LAD on the accuracy. The accuracy of multi-class LAD prognostic algorithm is improved by using the CDET as a feature reduction algorithm. However, this also support our hypothesis that the proposed methodology has stable performance over the other prediction models when the number of covariates is changing.

The expected deliverables of this prognostic methodology lead to the creation of an automated knowledge-base of the most frequently occurring faults in the bearings and their human interpretations. The automated knowledge-base can be employed to detect and predict the hidden phenomena for a given bearing condition. It is concluded from the obtained results that the proposed methodology is promising and tractable prognostic tool in the CBM. It allows the maintenance engineers to make reliable decisions without the need for an expert to examine and analyze the data. It also allows the system operator to easily predict the remaining life which is difficult to estimate by the conventional multiple failure modes prognostic methods, in the presence of many competing failure modes.

## 5.9 References

[1] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing,* vol. 20, pp. 1483-1510, 2006.

[2] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, "Intelligent fault diagnosis and prognosis for engineering systems, 2006," *Usa 454p Isbn,* vol. 13, pp. 978-0.

[3] A. Heng, A. C. Tan, J. Mathew, N. Montgomery, D. Banjevic, and A. K. Jardine, "Intelligent condition-based prediction of machinery reliability," *Mechanical Systems and Signal Processing,* vol. 23, pp. 1600-1614, 2009.

[4] A. Ragab, M.-S. Ouali, S. Yacout, and H. Osman, "Remaining useful life prediction using prognostic methodology based on logical analysis of data and Kaplan–Meier estimation," *Journal of Intelligent Manufacturing,* pp. 1-16, 2014.

[5] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing,* vol. 25, pp. 1803-1836, 2011.

[6] H. Putter, M. Fiocco, and R. Geskus, "Tutorial in biostatistics: competing risks and multi-state models," *Statistics in medicine,* vol. 26, pp. 2389-2430, 2007.

[7] D. Kleinbaum and M. Klein, "Survival Analysis: A self-learning text, 2005," ed: New York, Springer-Verlag, 2011.

[8] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data* vol. 360: John Wiley & Sons, 2011.

[9] R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson Jr, N. Flournoy, V. Farewell, and N. Breslow, "The analysis of failure times in the presence of competing risks," *Biometrics,* pp. 541-554, 1978.

[10] E. A. Elsayed, *Reliability engineering*: Wiley Publishing, 2012.

[11] J. Klein and M. Moeschberger, "Survival Analysis: Techniques for Censored and Truncated Data. 1997," *Spring-Verlag, New York,* pp. 276-278.

[12] M. Pintilie, "Analysing and interpreting competing risk data," *Statistics in medicine,* vol. 26, pp. 1360-1367, 2007.

[13] M. Pintilie, "An introduction to competing risks analysis," *Revista Española de Cardiología (English Edition),* vol. 64, pp. 599-605, 2011.

[14] E. Boros, P. L. Hammer, T. Ibaraki, A. Kogan, E. Mayoraz, and I. Muchnik, "An implementation of logical analysis of data," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 12, pp. 292-306, 2000.

[15] Y. Crama, P. L. Hammer, and T. Ibaraki, "Cause-effect relationships and partially defined Boolean functions," *Annals of Operations Research,* vol. 16, pp. 299-325, 1988.

[16] E. Bores, P. L. Hammer, T. Ibaraki, A. Kogan, E. Mayoraz, and I. Muchnik, "An implementation of logical analysis of data," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 12, pp. 292-306, 2000.

[17] M.-A. Mortada, S. Yacout, and A. Lakis, "Diagnosis of rotor bearings using logical analysis of data," *Journal of Quality in Maintenance Engineering,* vol. 17, pp. 371-397, 2011.

[18] T. O. Bonates, P. L. Hammer, and A. Kogan, "Maximum patterns in datasets," *Discrete Applied Mathematics,* vol. 156, pp. 846-861, 2008.

[19] C. M. Bishop, *Pattern recognition and machine learning* vol. 1: springer New York, 2006.

[20] M.-A. Mortada, S. Yacout, and A. Lakis, "Fault diagnosis in power transformers using multi-class logical analysis of data," *Journal of Intelligent Manufacturing,* pp. 1-11, 2013.

[21] J. F. A. Herrera and M. M. Subasi, "Logical Analysis of Multi-Class Data," 2013.

[22] S. Yacout, D. Salamanca, and M.-A. Mortada, "Tool and method for fault detection of devices by condition based maintenance," ed: Google Patents, 2011.

[23] E. A. Elsayed, "Mean residual life and optimal operating conditions for industrial furnace tubes," *Case Studies in Reliability and Maintenance*, pp. 497-515, 2003.

[24] H.-E. Kim, A. C. Tan, J. Mathew, and B.-K. Choi, "Bearing fault prognosis based on health state probability estimation," *Expert Systems with Applications,* vol. 39, pp. 5200-5213, 2012.

[25] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mechanical Systems and Signal Processing,* vol. 22, pp. 419-435, 2008.

[26] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*: John Wiley & Sons, 2011.

[27] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing,* vol. 23, pp. 724-739, 2009.

[28] P. Canada. Available: http://www.pruftechnik.ca/

[29] J. C. Goswami and A. K. Chan, *Fundamentals of wavelets: theory, algorithms, and applications* vol. 233: John Wiley & Sons, 2011.

[30] R. X. Gao and R. Yan, *Wavelets: Theory and Applications for manufacturing*: Springer, 2010.

[31] M. Misiti, Y. Misiti, G. Oppenheim, and J. Poggi, "Matlab user's guide: wavelet toolbox™ 4," *The Math Works Inc, Natick, Mass,* 2008.

[32] J. Qu and M. J. Zuo, "Support vector machine based data processing algorithm for wear degree classification of slurry pump systems," *Measurement,* vol. 43, pp. 781-791, 2010.

[33] B. Samanta, K. R. Al-Balushi, and S. A. Al-Araimi, "Artificial neural networks and genetic algorithm for bearing fault detection," *Soft Computing,* vol. 10, pp. 264-271, 2006.

[34] B. Yegnanarayana, *Artificial neural networks*: PHI Learning Pvt. Ltd., 2009.

[35] Z. Tian, L. Wong, and N. Safaei, "A neural network approach for remaining useful life prediction utilizing both failure and suspension histories," *Mechanical Systems and Signal Processing,* vol. 24, pp. 1542-1555, 2010.

[36] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2011.

[37] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," *Advances in neural information processing systems,* pp. 281-287, 1997.

[38] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann*, et al.*, "WEKA---Experiences with a Java Open-Source Project," *The Journal of Machine Learning Research,* vol. 11, pp. 2533-2541, 2010.

# CHAPTER 6 GENERAL DISCUSSION

This chapter discusses the contributions of the articles presented in Chapters 3, 4, and 5 towards accomplishing the specific objectives stated in the introduction of this thesis. The first specific objective is to implement the two-class LAD in CBM prognostic applications. It was achieved by developing two distinct prognostic methodologies; the first methodology is presented in Chapter 3 and the second one is presented in Chapter 4.

The first prognostic methodology discussed in Chapter 3 involves the exploitation of a historical CBM dataset (lifetime data and condition monitoring data) collected from systems working under different operating conditions. The condition monitoring data represent the operating conditions and condition indicators. The operating conditions are controllable, they are set to certain values, while the condition indicators are gathered as sensory information to reflect those operating conditions. The dataset consists of a set of observations collected from a set of systems, at/or immediately before the time to failure. Each system has only one observation consisting of the time to failure and a number of covariates that represent the operating conditions and condition indicators. The system that fails before a certain time $t_S$ (specified by the maintenance personnel) is called *Short-Life* (*SL*) system and the one that fails after that time is called *Long- Life* (*LL*) system. The time $t_S$ is set in this doctoral research to be the mean time to failure (MTTF) which is a feasible and practical choice and can be obtained easily by the maintenance personnel.

The two-class LAD is merged with KM estimatior as a non-parametric reliability estimation method that does not require any assuption about the distribution of the lifetime data. More specifically, LAD is used as an event-driven technique to deal with the covariates while KM is used as time-driven technique to deal with the lifetime data. Based on the lifetime data, KM is used to estimate the baseline survival curve. However, the estimated baseline survival curve does not reflect the effect of covariates on the monitored system. It reflects only the effect of the age on its health state. The two-class LAD is used as a machine learning classification technique to discover the hidden knowledge from the covariates in the form of generated patterns. An important advantage of the generated patterns is that they represent the high-degree of interactions among those correlated covariates.

A survival curve for each generated pattern is estimated by considering only the systems covered by that pattern. Consequently, the knowledge is extracted from the lifetime and condition monitoring data, in the form of non-parametric survival curves (the baseline survival curve and patterns' survival curves). The patterns' survival curves reflect the effect of covariates on the reliability of the monitored system.

Given an observation collected recently from a monitored system, the LAD decision rule along with these survival curves, are used to estimate a certain prognostic index; the RUL. The baseline curve of the monitored system is updated based on the covering patterns for that obervation. The updating step is carried out using one of two different models (*LAD Model 1 and LAD Model 2*). In both models, the survival curve of the monitored system is updated initially, by averaging the survival curves of the patterns that cover the first observation and the baseline curve. Then, the subsequent observations are used to update the survival curve of the system, by averaging the survival curves of the patterns that cover each observation and the former updated survival curve. The updated curve is then used to estimate the RUL of the monitored system.

In *LAD Model 1*, the survival curve for each generated pattern has the same weight as the baseline survival curve. Therefore, two situations can happen most probably. In the first situation, the resulting updated survival curve can have a pessimistic explanatory style or outlook, if the observation is covered by a large number of *SL* patterns. Pessimistic outlook means that the updated survival curve can be decreased significantly than it should be, due to the coverage of the *SL* patterns. Having a pessimistic outlook can be interpreted physically as that the monitored system attributes bad symptoms that lead to a catastrophic failure. This can result in an underestimated RUL for the monitored system. In the second situation, the updated survival curve can have an optimistic outlook if the observation is covered by a large number of *LL* patterns. The resulting updated curve interprets this situation as being best (optimized). In this case, the updated survival curve can be increased more than it should be. The optimistic outlook means that the system attributes good symptoms, which is interpreted as surviving more. This can lead to an overestimation of the system's RUL.

The extent in which the above two situations are evaluated as something bad or something good in the monitored system may not be fully comprehended in an optimum way. As a consequence of the two situations, the survival information of the monitored system will not be explored properly.

In *LAD Model 2*, the baseline survival curve has a greater contribution than the survival curves of the patterns. In other words, the contribution of each pattern's survival curve in the updating model is less than that of the baseline curve. This can help to decrease the amount of pessimism or optimism in the updated survival curve.

The amount of decrease or increase in the updated survival curve can be illustrated in the following.

For *LAD Model 1*, at the time which the first observation is collected from the monitored system, the averaging formula is given as:

$$S_{Model\,1}(t) = \frac{[\sum_{j=1}^{n} S_{p_j}(t) + S_b(t)\,]}{(n+1)} = \frac{S_{p_1}(t)}{(n+1)} + \frac{S_{p_2}(t)}{(n+1)} + \cdots + \frac{S_{p_n}(t)}{(n+1)} + \frac{S_b(t)}{(n+1)} \qquad (6.1)$$

For *LAD Model 2*, the averaging formula is given as:

$$S_{Model\,2}(t) = \frac{\frac{[\sum_{j=1}^{n} S_{p_j}(t)\,]}{n} + S_b(t)}{2} = \frac{S_{p_1}(t)}{2n} + \frac{S_{p_2}(t)}{2n} + \cdots \frac{S_{p_n}(t)}{2n} + \frac{S_b(t)}{2} \qquad (6.2)$$

Mathematically, we can prove that *LAD Model 1* gives more pessimistic (optimistic) survival curve than *LAD Model 2*, if the collected observation is covered by only *SL* (*LL*) patterns. Therefore, we need to prove the following:

1- $S_{Model\,2}(t) > S_{Model\,1}(t)$ if $S_{p_j}(t) < S_b(t)$ for all $j = 1, \ldots n$
2- $S_{Model\,2}(t) < S_{Model\,1}(t)$ if $S_{p_j}(t) > S_b(t)$ for all $j = 1, \ldots n$

This can be proved if we subtract equation (6.1) from equation (6.2). This yields,

$$\begin{aligned} S_{Model\,2}(t) - S_{Model\,1}(t) &= \left[\tfrac{1}{2n} - \tfrac{1}{n+1}\right] S_{p_1}(t) + \cdots \left[\tfrac{1}{2n} - \tfrac{1}{n+1}\right] S_{p_n}(t) + \left[\tfrac{1}{2} - \tfrac{1}{n+1}\right] S_b(t) \\ &= \left[\frac{1-n}{2n(n+1)}\right] \left[S_{p_1}(t) + S_{p_2}(t) + \cdots S_{p_n}(t)\right] + \left[\frac{n-1}{2(n+1)}\right] S_b(t) \qquad (6.3) \end{aligned}$$

We notice from equation (6.3) that the term $\left[\frac{n-1}{2(n+1)}\right]$ which is multiplied by $S_b(t)$ is equal to $-n$ times the term $\left[\frac{1-n}{2n(n+1)}\right]$ which is multiplied by the sum of all survival functions of the covering patterns. It means that equation (6.3) can have one of possible three values, given as:

$$\left[\frac{1-n}{2n(n+1)}\right] \left[S_{p_1}(t) + S_{p_2}(t) + \cdots S_{p_n}(t)\right] + \left[\frac{n-1}{2(n+1)}\right] S_b(t) = \begin{cases} = 0 & \text{if } \; n = 1 \\ > 0 & \text{if } \; S_{p_j}(t) < S_b(t) \quad \forall\; j \\ < 0 & \text{if } \; S_{p_j}(t) > S_b(t) \quad \forall\; j \end{cases}$$

For equation (6.3), having a value of zero means that the two models are identical. This is the case when the collected observation is covered by only one pattern. Having a value that is greater (less) than zero means that *LAD Model 1* is more pessimistic (optimistic) than *LAD Model 2*. It means that *LAD Model 2* is more informative and have more interpretation than *LAD Model 1*. However, we can only prove that when the collected observation is covered by only *SL* (or *LL*) patterns. This proof will not be valid if the observation is covered by a mixture of *SL* and *LL* patterns. Therefore, it is necessary to adapt the updating model by assigning a weight for each of the patterns' survival curves. This may be helpful to decrease such kind of pessimistic (or optimistic) behaviour.

Based on the two updating models, the proposed LAD methodology was compared to the PHM prediction model through Friedman test. The obtained results from this test conclude that the LAD prognostic methodology gives more accurrate prediction for the RUL than the PHM prediction model. Moreover, the results conclude that there is a significant difference between the performance of LAD methodology and that of the PHM, in particular when *LAD Model 2* is used.

Chapter 4 presents an enhanced prognostic methodology that proposes three modifications to the the prognostic methodology presented in Chapter 3. The modifications are summarized in the following:

1- This enhanced methodology exploits all the CBM data (both normal and failure observations) collected from the systems during their lifespans. This means that all the bulk of information embedded in each system's life are not ignored.

2- A pattern selection procedure is used in order to select the most significant patterns form all the generated patterns. This is because the number of generated pattern from the lifespan data is large and some of them are redundant and cover small number of obervations.

3- A weight that reflects the coverage of each pattern, thus its importance, is assigned to its survival curve. Accordingly, a new updating model (*LAD Model 3*) that considers the weights of the patterns, is proposed. Our hypothesis states that the new proposed model can decrease the amount of pessimism or optimism in the updating curve, significantly.

In Chapter 4, a comparison was conducted between the two updating models (*LAD Model 1* and *LAD Model 2)*, and the proposed model (*LAD Model 3*). It is concluded from the computational results that there is a significant difference between the performance of the two models (*LAD Model 2* and *LAD Model 3*) and that of *LAD Model 1*. This difference is in favor of the formers. From the obtained results, it is also noticed that the pattern selection procedure has an effect on the performance of the enhanced LAD prognostic methodology, because it removes the redundant patterns which have small weights. The comparisons between the three updating models, using Friedman test, show that the proposed methodology yields an improved accuracy for the RUL estimation, when it considers the selected patterns and their weights. To illustrate that the proposed methodology in Chapter 4 exploits all the CBM data, it was compared to two of the most common machine learning regression techniques; the ANN and SVR. The results show that there are significant differences between the performance of *LAD Model 2* and *LAD Model 3,* and those of *ANN* and *SVR.* It is also concluded from the comparisons that *LAD Model 3* and *LAD Model 2* are the best, *SVR* and *LAD Model 1* come in second, and *ANN Model* is the worst. The results support our hypothesis which states that the proposed methodology exploits effectively all the CBM data, and gives an accurate prediction for the RUL, in particular when *LAD Model 3* is used in the updating phase.

The second objective stated in the introduction of this thesis is accomplished through the application of multi-class LAD to construct the third prognostic methodology, which is presented in Chapter 5. The methodology considers the failure of the complex systems that are subjected to interactions and competitions between different failure modes. Two limitations of the current multiple failure modes prognostic models were addressed in the chapter. They were stated in the following. First, many of those models are parametric, they assign a certain failure distribution function for the lifetime data of each failure mode. This needs a lot of experience and knowledge about the application at hand. Second, Such models assume that the failure distribution function for each failure mode is estimated by considering the other competing modes as censored categories. The limitations were addressed in that prognostic methodology, by estimating a non-parametric cumulative incidence function for each failure mode, based on its historical lifetime data. Each estimated function takes into account the competition between the different failure modes. The multi-class LAD approach is used on the other hand to generate a set of patterns from the historical condition monitoring data that reflect the interactions between the covariates in those failure modes. Accordingly, the generated patterns represent the interactions between the covariates of such historical data, and they differentiate between the observations in each failure mode. Each generated pattern is allowed to cover a large proportion of observations in the corresponding class (observations in a given failure mode), and small proportion from the other classes. The knowledge is extracted in the form of multi-class LAD decision rule (scoring function) and a set of CIF curves. Given a new observation collected from a monitored system, its survival curve is updated based on the score obtained by that decision rule, then its RUL is caculated.

The application discussed in Chapter 5 involves the RUL prediction of the bearings in rotating machinery, by using vibration data collected from a real application in the industry. The data contain a number of time waveform signals collected from a set of bearings subjected to one of many failure modes. Each bearing starts running from its new state until the occurrence of complete failure. The signals are processed and a set of features are extracted. A set of significant features are selected from all the extracted ones. The patterns are generated based on the given features to provide physical interpretations for the failure modes, thus not only help as diagnostic information but also aid in updating the prognostic measures to avoid the occurrence of bearing's failure. The proposed methodology in Chapter 5 was validated based on two experiments. The first experiment involves all the extracted features, while the second involves the set of selected features. In those experiments, the proposed LAD methodology performed well against the ANN and SVR and gave a stable performance. The results show that it is promising and can accurately predict the RUL of a monitored system in the presence of multiple failure modes.

The developed prognostic methodologies in this doctoral research are validated and justified by the following:

1- The resulting performance in terms of the difference between the estimated and actual RUL of the monitored systems.
2- The advantage of LAD which possesses some important merits over the other machine learning techniques and its beneficial role that supports the CBM decision makers with an adequate and updated prognostic knowledge. Merging LAD to non-parametric estimation methods can deal properly with the prognostic problems mentioned in the introduction of this thesis.
3- The implementation of the prognostic methodologies in different practical situations, in terms of the type of analysis of failure modes being addressed, the nature of the CBM data (failure data or lifespan data), and the independency of the application at hand. The work presented in this thesis can be employed to reduce industry's dependence on experts.

# CONCLUSION AND FUTURE WORK

Major advances have been achieved in machine diagnostics which is a subject of considerable attention in the condition-based maintenance (CBM) domain. However, the CBM prognostics has not enjoyed the same attention. A particular prognostic task is the prediction of system's remaining useful life (RUL) that attracts a significant number of researchers, as the monitored systems become complex and critical. The RUL is defined as the time left before the occurrence of complete failure in the system. The need for accurate prognostic techniques has become urgent to maximize the reliability (survival probability) of such critical systems, hence survive for a long period of time.

The estimated survival function using either parametric or non-parametric methods, is used to predict the failure time of the monitored system, which in turn allows the prediction of its RUL. This kind of prediction is called reliability-based prognosis. However, the prediction of the RUL may not be accurrate if the operating conditions of the monitored system are not involved. The use of condition monitoring data has a significant effect on the performance of the RUL prediction. Therefore, a particular challenge in CBM prognostic is to estimate the RUL of a monitored system working under varying operating conditions. This is because the relationship between condition measurements and the RUL in many situations is complicated and not fully understood.

Many statistical prognostic methods are applied in CBM to estimate the reliability function of the monitored systems that are working under different operating conditions. Proportional Hazards Model (PHM) as an example is one of the common statistical CBM prognostic methods. One limitation for such statistical methods is that they assume a certain probability distribution for the lifetime data, in order to estimate the survival function.

In this doctoral research, this limitation was addressed by using a non-parametric method called Kaplan-Meier (KM) to estimate the survival function. Another limitation of those statistical methods is that they could not deal with the covariates (the operating conditions and condition indicators) that are highly correlated or time-varying. Other prognostic methods are mainly dependent on accumulating knowledge or experience about the targeted application.

The subject of this doctoral research to apply a relatively new knowledge discovery approach called Logical Analysis of Data (LAD) in the field of CBM prognostics. LAD is a pattern-based machine learning technique that originated from two disparate fields; Boolean functions and combinatorial optimization. The objective is use LAD to exploit the CBM databases in order to discover the hidden knowledge. The data are collected from a set of historical systems, containing condition indicators, are exploited to generate patterns that relate those indicators to the RUL of the monitored systems.

LAD as a diagnostic technique has two advantages over many machine learning techniques. LAD is not based on any statistical analysis. Therefore, it is capable of dealing with covariates that are highly correlated or time-varying, without the need to satisfy any statistical assumptions. One of the most important advantages of the LAD approach when compared to the other machine learning techniques in CBM is its transparency (interpretability) and knowledge preservation. This allows interpreting the discovered knowledge in the CBM databases (in the form of patterns) to a beneficial physical meaning. The obtained patterns are powerful tools for constructing a decision model that is useful for predicting newly unseen data. The discovered knowledge can be preserved for the future use by the maintenance personnel.

The original contribution of this thesis is the introduction of innovative prognostic methodologies based on LAD, to the domain of CBM. In the first one, the KM and LAD are merged together to build a novel prognostic methodology. Given a set of observations collected from historical systems, the knowledge is extracted from the lifetime and the condition monitoring data, in the form of non-parametric survival curves. Based on the diagnostic information obtained from the LAD decision model, these curves (which reflect the effect of operating conditions) were used to estimate certain prognostic indices such as the RUL of the monitored system. The first methodology was compared to the PHM prediction model. The results showed that the LAD methodology provides a more accurate RUL prediction than the compared model.

The second methodology involved three modifications to the first methodology, which resulted in an improved performance and achieved higher accuracy. In addition, it achieved the best overall accuracy when compared to other popular machine learning algorithms.

The third methodology in this doctoral research addressed another challenge which is the RUL prediction in the presence of multiple competing failure modes. The multi-class LAD is merged with a set of non-parametric functions called cumulative incidence functions (CIFs), to construct a prognostic methodology for the systems that fail due to multiple failure modes. A well known application was employed from the field of rotating machinery, to illustrate the effectiveness of the methodology. In this application, the vibration data are collected from a set of bearings. The obtained patterns have obvious physical meaning and the failure modes are linked back to the measured symptoms of the bearing. A comparison was conducted to validate the methodology. It is concluded from the obtained results that it outperforms the compared techniques.

It is concluded that the three methodologies devised in this thesis were demonstrated higher performance than the other compared methods. As a final conclusion from obtained results in this doctoral research, LAD is a promising knowledge discovery approach which can be easily adapted to a multitude of CBM prognostic applications.

In this doctoral research, we have identified a limitation of the proposed methodologies. All of these methodologies are based on LAD which is a combinatorial optimization-based approach. LAD uses Mixed Integer and Linear Programming (MILP) formulation to generate the patterns from the training dataset. The MILP formulation is solved iteratively until all the observations in the training dataset are covered by the generated patterns. This makes the pattern generation procedure computationally demanding. This process is time consuming in case of datasets with large number of observations. As such, we are planning to focus on the development of new pattern generation procedures that would help applying LAD as a CBM decision model in online applications.

We also identified another limitation on the applicability of LAD as unsupervised learning approach. The current LAD approach depends on the existence of labeled data, in order to train the decision model. Our future plan is to apply LAD as unsupervised learning approach, either by modifying its structure or by linking it a suitable and compatible clustering technique. This could be more feasible and may help to separate the data automatically without setting any parameters.

As a future work, we are planing to create an automated knowledgebase for the purpose of building a tele-maintenance system. Building this knowledgebase takes months or even years when relying on the expertise of the maintenance engineers. The developed prognostic methodologies in this doctoral research will be utilized in addition to the diagnostic capability of LAD, as the cornerstones of this automated knowledgebase. We already started the work in this direction in our lab at École Polytechnique de Montréal, by building a prototype system to collect the vibration data from rotating machines. The objective is the creation of an automated and updated knowledge for the most frequently occurring faults in rotor bearings and their human interpretations.

The automated diagnostic/prognostic knowledgebase will be exploited to build an online tele-maintenance system, in which the maintenance personnel can detect the faults and predict the RUL remotely, without the need to visit the system continuously. This can be helpful to reduce the maintenance cost significantly since the maintenance personnel acts only when a necessary action should be taken.

# REFERENCES


[1] L. Ma, "Condition monitoring in engineering asset management," 2007.

[2] R. Kothamasu, S. H. Huang, and W. H. VerDuin, "System health monitoring and prognostics—a review of current paradigms and practices," *The International Journal of Advanced Manufacturing Technology,* vol. 28, pp. 1012-1024, 2006.

[3] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing,* vol. 20, pp. 1483-1510, 2006.

[4] J. D. Campbell and A. K. S. Jardine, *Maintenance excellence: Optimizing equipment life-cycle decisions* vol. 135: CRC, 2001.

[5] R. K. Mobley, *An introduction to predictive maintenance*: Butterworth-Heinemann, 2002.

[6] T. Wireman, *Benchmarking best practices in maintenance management*: Industrial Press Inc., 2004.

[7] A. Garg and S. Deshmukh, "Maintenance management: literature review and directions," *Journal of Quality in Maintenance Engineering,* vol. 12, pp. 205-238, 2006.

[8] A. C. Márquez, *The maintenance management framework: models and methods for complex systems maintenance*: Springer Verlag, 2007.

[9] A. K. Jardine and A. H. Tsang, *Maintenance, replacement, and reliability: theory and applications*: CRC press, 2013.

[10] M. A. Mortada, T. Carroll, S. Yacout, and A. Lakis, "Rogue components: their effect and control using logical analysis of data," *Journal of Intelligent Manufacturing,* pp. 1-14, 2009.

[11] A. M. Smith and G. R. Hinchcliffe, *RCM--Gateway to World Class Maintenance*: Butterworth-Heinemann, 2003.

[12] M. Pecht, *Prognostics and health management of electronics*: Wiley Online Library, 2008.

[13] A. Jardine, T. Joseph, and D. Banjevic, "Optimizing condition-based maintenance decisions for equipment subject to vibration monitoring," *Journal of Quality in Maintenance Engineering,* vol. 5, pp. 192-202, 1999.

[14] A. Tan, A. S. Y. Heng, and J. Mathew, "Condition-based prognosis of machine health," 2009, pp. 1-10.

[15] A. Bayoumi, N. Goodman, R. Shah, L. Eisner, L. Grant, J. Keller, *et al.*, "Conditioned-based maintenance at USC—Part IV: Examination and cost-benefit analysis of the CBM Process," 2008.

[16] S. Das, R. Hall, A. Patel, S. McNamara, and J. Todd, "An open architecture for enabling CBM/PHM capabilities in ground vehicles," 2012, pp. 1-8.

[17] K. R. Wheeler, T. Kurtoglu, and S. D. Poll, "A Survey of Health Management User Objectives in Aerospace Systems Related to Diagnostic and Prognostic Metrics," *Int. J. Prog. & Health. Mangt*, 2010.

[18] IAEA, "Implementation Strategies and Tools for Condition Based Maintenance at Nuclear Power Plants," VIENNAMay 2007 2007.

[19] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *Journal of Intelligent Manufacturing*, pp. 1-11, 2009.

[20] G. J. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent fault diagnosis and prognosis for engineering systems*: Wiley Online Library, 2006.

[21] Z. Tian, L. Wong, and N. Safaei, "A neural network approach for remaining useful life prediction utilizing both failure and suspension histories," *Mechanical Systems and Signal Processing*, vol. 24, pp. 1542-1555, 2010.

[22] H. E. Kim, A. C. C. Tan, J. Mathew, and B. K. Choi, "Bearing fault prognosis based on health state probability estimation," *Expert Systems with Applications*, 2011.

[23] S. Alexe, E. Blackstone, P. L. Hammer, H. Ishwaran, M. S. Lauer, and C. E. Pothier Snader, "Coronary risk prediction by logical analysis of data," *Annals of Operations Research*, vol. 119, pp. 15-42, 2003.

[24] G. Alexe, S. Alexe, T. O. Bonates, and A. Kogan, "Logical analysis of data–the vision of Peter L. Hammer," *Annals of Mathematics and Artificial Intelligence*, vol. 49, pp. 265-312, 2007.

[25] L. P. Kronek and A. Reddy, "Logical analysis of survival data: prognostic survival models by detecting high-degree interactions in right-censored data," *Bioinformatics*, vol. 24, pp. i248-i253, 2008.

[26] E. Bores, P. L. Hammer, T. Ibaraki, A. Kogan, E. Mayoraz, and I. Muchnik, "An implementation of logical analysis of data," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 12, pp. 292-306, 2000.

[27] D. Salamanca and S. Yacout, "Condition based maintenance with logical analysis of data," *7e Congrès International de génie industriel,* 2007.

[28] E. A. Elsayed, *Reliability engineering*: Wiley Publishing, 2012.

[29] M. Schwabacher and K. Goebel, "A survey of artificial intelligence for prognostics," 2007, pp. 107-114.

[30] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing,* vol. 25, pp. 1803-1836, 2011.

[31] M.-A. Mortada, S. Yacout, and A. Lakis, "Diagnosis of rotor bearings using logical analysis of data," *Journal of Quality in Maintenance Engineering,* vol. 17, pp. 371-397, 2011.

[32] M.-A. Mortada, S. Yacout, and A. Lakis, "Fault diagnosis in power transformers using multi-class logical analysis of data," *Journal of Intelligent Manufacturing,* pp. 1-11, 2013.

[33] J. H. Williams, A. Davies, and P. R. Drake, *Condition-based maintenance and machine diagnostics*: Kluwer Academic Publishers, 1994.

[34] A. Heng, A. C. C. Tan, J. Mathew, N. Montgomery, D. Banjevic, and A. K. S. Jardine, "Intelligent condition-based prediction of machinery reliability," *Mechanical Systems and Signal Processing,* vol. 23, pp. 1600-1614, 2009.

[35] A. Labib, "Computerised maintenance management systems," *Complex System Maintenance Handbook*, pp. 417-435, 2008.

[36] L. Adjengue, S. Yacout, and O. Ilk, "Parameters Estimation for Condition Based Maintenance with Uncorrelated and Correlated Observations," *Quality Engineering*, vol. 19, pp. 197-206, 2007.

[37] M. Bengtsson, *Condition Based Maintenance Systems: An Investigation of Technical Constituents and Organizational Aspects*: Citeseer, 2004.

[38] M. Bengtsson, "On condition based maintenance and its implementation in industrial settings," 2007.

[39] M. G. Thurston, "An open standard for web-based condition-based maintenance systems," 2001, pp. 401-415.

[40] M.-D. Gauvain, R. Gouriveau, N. Zerhouni, C. Varnier, N. Piat, and M. Hessabi, "Defining and implementing a distributed and reconfigurable information system for prognostics," in *Prognostics and System Health Management Conference (PHM-Shenzhen), 2011*, 2011, pp. 1-6.

[41] M. Hamada, "Using degradation data to assess reliability," *Quality Engineering,* vol. 17, pp. 615-620, 2005.

[42] N. Jammu and P. Kankar, "A Review on Prognosis of Rolling Element Bearings," *International Journal of Engineering Science*, vol. 3, 2011.

[43] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2011.

[44] A. Prajapati, J. Bechtel, and S. Ganesan, "Condition based maintenance: a survey," *Journal of Quality in Maintenance Engineering,* vol. 18, pp. 384-400, 2012.

[45] C. McCollin and S. Coleman, "Historical Published Maintenance Data: What Can It Tell Us About Reliability Modelling?," *Quality and Reliability Engineering International,* vol. 30, pp. 381-395, 2014.

[46] D. Barraza-Barraza, J. Limón-Robles, N. L. Monterrey, and M. G. Beruvides, "Opportunities and Challenges in Condition-Based Maintenance Research."

[47] P. Giudici, *Applied data mining: statistical methods for business and industry*: John Wiley & Sons, 2005.

[48] C. M. Bishop and SpringerLink, *Pattern recognition and machine learning* vol. 4: springer New York, 2006.

[49] F. Gorunescu, *Data Mining: Concepts, models and techniques* vol. 12: Springer, 2011.

[50] S. Ebersbach, "Artificial intelligent system for integrated wear debris and vibration analysis in machine condition monitoring," James Cook University, 2007.

[51] J. S. Mitchell, "From Vibration Measurements to Condition-Based Maintenance," *Sound and Vibration,* vol. 41, p. 62, 2007.

[52] S. Telford, M. I. Mazhar, and I. Howard, "Condition Based Maintenance (CBM) in the Oil and Gas Industry: An Overview of Methods and Techniques."

[53] A. Witney, Y. F. Li, J. Wang, M. Wang, J. DeLuccia, and C. Laird, "Electrochemical fatigue sensor response to Ti–6 wt% Al–4 wt% V and 4130 steel," *Philosophical Magazine,* vol. 84, pp. 331-349, 2004.

[54] F. Alloys, N. F. Alloys, P. Metal, T. Leaders, A. Testers, A. Ovens*, et al.*, "RF Proximity Sensors-Characteristics, Operation and Application in Materials Recycling," *Materials World,* vol. 10, pp. 21-22, 2002.

[55] F. M. Discenzo, D. Chung, and K. A. Loparo, "Pump condition monitoring using self-powered wireless sensors," *Sound and Vibration,* vol. 40, pp. 12-15, 2006.

[56] A. Tiwari, F. Lewis, and S. S. Ge, "Wireless sensor network for machine condition based maintenance," 2004, pp. 461-467 Vol. 1.

[57] K. Srinivasan, M. Ndoh, H. Nie, C. Xia, D. Ingraham, and K. Kaluri, "Wireless technologies for condition-based maintenance (CBM) in petroleum plants," *Lecture Notes in Computer Science, Springer-Verlag, v3560,* 2005.

[58] R. Pallas-Areny and J. G. Webster, "Sensors & signal conditioning," *Recherche,* vol. 67, p. 02, 2000.

[59] W. Kester and A. Devices, *Practical design techniques for sensor signal conditioning*: Analog Devices, 1999.

[60] M. Abbas, A. A. Ferri, M. E. Orchard, and G. J. Vachtsevanos, "An intelligent diagnostic/prognostic framework for automotive electrical systems," 2007, pp. 352-357.

[61] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification. A Wiley-Interscience Publication," ed: John Wiley & Sons, Inc., 2001.

[62] I. T. Jolliffe and MyiLibrary, *Principal component analysis* vol. 2: Wiley Online Library, 2002.

[63] M. Welling, "Fisher linear discriminant analysis," *Department of Computer Science, University of Toronto,* 2005.

[64] J. V. Stone, *Independent component analysis: a tutorial introduction*: The MIT Press, 2004.

[65] C. Ellwein, S. Danaher, and U. Jager, "Identifying regions of interest in spectra for classification purposes," *Mechanical Systems and Signal Processing,* vol. 16, pp. 211-222, 2002.

[66] T. Heger and M. Pandit, "Optical wear assessment system for grinding tools," *Journal of Electronic Imaging,* vol. 13, p. 450, 2004.

[67] R. B. Randall, *Vibration-based condition monitoring: industrial, aerospace and automotive applications*: John Wiley & Sons, 2011.

[68] M. Lebold, K. McClintic, R. Campbell, C. Byington, and K. Maynard, "Review of vibration analysis methods for gearbox diagnostics and prognostics," 2000, pp. 623-634.

[69] M. A. Mortada, S. Yacout, and A. Lakis, "Fault Diagnosis of Power Transformers Using Logical Analysis of Data," *APPLICABILITY AND INTERPRETABILITY OF LOGICAL ANALYSIS OF DATA IN CONDITION BASED MAINTENANCE,* p. 74, 2010.

[70] M. Yang, "Early Fault Detection Scheme and Optimized CBM Strategy for Gear Transmission System Operating under Varying Loads," University of Toronto, 2011.

[71] R. X. Gao and R. Yan, *Wavelets: Theory and Applications for Manufacturing*: Springer Verlag, 2010.

[72] H. A. Gaberson, "The use of wavelets for analyzing transient machinery vibration," *Sound and Vibration,* vol. 36, pp. 12-17, 2002.

[73] A. Lakis, "Rotating machinery fault diagnosis using time-frequency methods," 2007.

[74] P. S. R. Diniz, E. A. B. Da Silva, and S. L. Netto, *Digital signal processing: system analysis and design*: Cambridge Univ Pr, 2010.

[75] J. C. Goswami and A. K. Chan, *Fundamentals of wavelets: theory, algorithms, and applications* vol. 233: John Wiley & Sons, 2011.

[76] R. Yan and R. X. Gao, "An efficient approach to machine health diagnosis based on harmonic wavelet packet transform," *Robotics and Computer-Integrated Manufacturing,* vol. 21, pp. 291-301, 2005.

[77] Y. Lei, Z. He, and Y. Zi, "A new approach to intelligent fault diagnosis of rotating machinery," *Expert Systems with Applications,* vol. 35, pp. 1593-1600, 2008.

[78] B. S. Yang, T. Han, and W. W. Hwang, "Fault diagnosis of rotating machinery based on multi-class support vector machines," *Journal of mechanical science and technology,* vol. 19, pp. 846-859, 2005.

[79] S. Knerr, L. Personnaz, G. Dreyfus, J. Fogelman, A. Agresti, M. Ajiz*, et al.*, "Single-layer learning revisited: A stepwise procedure for building and training a neural network," *Optimization Methods and Software,* vol. 1, pp. 23-34, 1990.

[80] B. Samanta, K. Al-Balushi, and S. Al-Araimi, "Artificial neural networks and genetic algorithm for bearing fault detection," *Soft Computing-A Fusion of Foundations, Methodologies and Applications,* vol. 10, pp. 264-271, 2006.

[81] T. Han and B. S. Yang, "Development of an e-maintenance system integrating advanced techniques," *Computers in Industry,* vol. 57, pp. 569-580, 2006.

[82] L. Zhang, L. B Jack, and A. K. Nandi, "Fault detection using genetic programming," *Mechanical Systems and Signal Processing,* vol. 19, pp. 271-289, 2005.

[83] J. Qu and M. J. Zuo, "Support vector machine based data processing algorithm for wear degree classification of slurry pump systems," *Measurement,* vol. 43, pp. 781-791, 2010.

[84] A. Widodo, B. S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors," *Expert Systems with Applications,* vol. 32, pp. 299-312, 2007.

[85] Y. Lei, Z. He, and Y. Zi, "Application of an intelligent classification method to mechanical fault diagnosis," *Expert Systems with Applications,* vol. 36, pp. 9941-9948, 2009.

[86] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mechanical Systems and Signal Processing,* vol. 21, pp. 688-705, 2007.

[87] K. Holmberg, A. Helle, and J. Halme, "Prognostics for industrial machinery availability," 2005.

[88] N. M. Vichare and M. G. Pecht, "Prognostics and health management of electronics," *Components and Packaging Technologies, IEEE Transactions on,* vol. 29, pp. 222-229, 2006.

[89] D. Tobon-Mejia, K. Medjaher, and N. Zerhouni, "CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks," *Mechanical Systems and Signal Processing,* vol. 28, pp. 167-182, 2012.

[90] J. Klein and M. Moeschberger, "Survival Analysis: Techniques for Censored and Truncated Data. 1997," *Spring-Verlag, New York,* pp. 276-278.

[91] D. G. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text, Third Edition*: Springer, 2011.

[92] M. Hamada, *Bayesian reliability*: Springer Verlag, 2008.

[93] E. A. Elsayed, "Mean residual life and optimal operating conditions for industrial furnace tubes," *Case Studies in Reliability and Maintenance,* pp. 497-515, 2003.

[94] D. Banjevic and A. Jardine, "Remaining useful life in condition based maintenance: Is it useful?," *Modelling in Industrial Maintenance and Reliability,* p. 7, 2007.

[95] A. B. Cantor, *SAS survival analysis techniques for medical research*: SAS publishing, 2003.

[96] Q. Zhang, C. Hua, and G. Xu, "A mixture Weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data," *Mechanical Systems and Signal Processing,* vol. 43, pp. 103-112, 2014.

[97] R. L. Prentice, J. D. Kalbfleisch, A. V. Peterson Jr, N. Flournoy, V. Farewell, and N. Breslow, "The analysis of failure times in the presence of competing risks," *Biometrics,* pp. 541-554, 1978.

[98] J. D. Kalbfleisch and R. L. Prentice, *The statistical analysis of failure time data* vol. 360: John Wiley & Sons, 2011.

[99] H. A. Awad, M. M. Gomaa, and A. R. Anwar, "Discrete event control and fault detection and localization of a robotic system," 2009, pp. 86-93.

[100] R. David, H. Alla, and I. NetLibrary, "Discrete, continuous, and hybrid Petri nets," 2005.

[101] T. Murata, "Petri nets: Properties, analysis and applications," *Proceedings of the IEEE,* vol. 77, pp. 541-580, 1989.

[102] M. Schwabacher, "A survey of data-driven prognostics," 2005.

[103] M. Krupa, "METHODS OF TECHNICAL PROGNOSIS-REVIEW."

[104] J. B. Coble, P. Ramuhalli, L. J. Bond, J. Hines, and B. Upadhyaya, *Prognostics and health management in nuclear power plants: a review of technologies and applications*: Pacific Northwest National Laboratory, 2012.

[105] F. Ahmadzadeh and J. Lundberg, "Remaining useful life estimation: review," *International Journal of System Assurance Engineering and Management,* pp. 1-14.

[106] C. Xiongzi, Y. Jinsong, T. Diyin, and W. Yingxun, "Remaining useful life prognostic estimation for aircraft subsystems or components: A review," in *Electronic Measurement & Instruments (ICEMI), 2011 10th International Conference on*, 2011, pp. 94-98.

[107] A. S. Sait and Y. I. Sharaf-Eldeen, "A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics," in *Rotating Machinery, Structural Health Monitoring, Shock and Vibration, Volume 5*, ed: Springer, 2011, pp. 307-324.

[108] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing,* vol. 23, pp. 724-739, 2009.

[109] M. Luo, D. Wang, M. Pham, C. Low, J. Zhang, D. Zhang, *et al.*, "Model-based fault diagnosis/prognosis for wheeled mobile robots: A review," 2005, p. 6 pp.

[110] S. Marble and B. P. Morton, "Predicting the remaining life of propulsion system bearings," 2006, p. 8 pp.

[111] A. Saxena, J. R. Celaya, I. Roychoudhury, S. Saha, B. Saha, and K. Goebel, "Designing Data-Driven Battery Prognostic Approaches for Variable Loading Profiles: Some Lessons Learned."

[112] K. Goebel, B. Saha, A. Saxena, J. Celaya, and J. Christophersen, "Prognostics in battery health management," *Instrumentation & Measurement Magazine, IEEE,* vol. 11, pp. 33-40, 2008.

[113] C. Sankavaram, B. Pattipati, A. Kodali, K. Pattipati, M. Azam, S. Kumar, *et al.*, "Model-based and data-driven prognosis of automotive and electronic systems," 2009, pp. 96-101.

[114] L. Qipeng, Y. Xiaoling, and F. Quanke, "Fault diagnosis using wavelet neural networks," *Neural processing letters,* vol. 18, pp. 115-123, 2003.

[115] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *Industrial Electronics, IEEE Transactions on,* vol. 51, pp. 694-700, 2004.

[116] Z. Tian and M. J. Zuo, "Health condition prediction of gears using a recurrent neural network approach," *Reliability, IEEE Transactions on,* vol. 59, pp. 700-705, 2010.

[117] C. S. Byington, M. Watson, and D. Edwards, "Data-driven neural network methodology to remaining life predictions for aircraft actuator components," 2004, pp. 3581-3589 Vol. 6.

[118] J. Liu, D. Djurdjanovic, J. Ni, N. Casoetto, and J. Lee, "Similarity based method for manufacturing process performance prediction and diagnosis," *Computers in Industry,* vol. 58, pp. 558-566, 2007.

[119] E. Zio and F. Di Maio, "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system," *Reliability Engineering & System Safety,* vol. 95, pp. 49-57, 2010.

[120] O. P. Yadav, N. Choudhary, and C. Bilen, "Complex system reliability estimation methodology in the absence of failure data," *Quality and reliability engineering international,* vol. 24, pp. 745-764, 2008.

[121] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mechanical Systems and Signal Processing,* vol. 24, pp. 1161-1171, 2010.

[122] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning* vol. 1: Springer Series in Statistics, 2001.

[123] A. H. Tsang, W. Yeung, A. K. Jardine, and B. P. Leung, "Data management for CBM optimization," *Journal of Quality in Maintenance Engineering,* vol. 12, pp. 37-51, 2006.

[124] N. Montgomery, T. Lindquist, M.-A. Garnero, R. Chevalier, and A. Jardine, "Reliability functions and optimal decisions using condition data for EDF primary pumps," in *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, 2006, pp. 1-6.

[125] M. Tail, S. Yacout, and M. Balazinski, "Replacement time of a cutting tool subject to variable speed," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* vol. 224, pp. 373-383, 2010.

[126] Z. Tian, D. Lin, and B. Wu, "Condition based maintenance optimization considering multiple objectives," *Journal of Intelligent Manufacturing,* pp. 1-8, 2009.

[127] Z. Tian and H. Liao, "Condition based maintenance optimization for multi-component systems using proportional hazards model," *Reliability Engineering & System Safety,* 2011.

[128] F. Tianle and Z. Jianmin, "SVM-PHM: A Novel Method for Remaining Useful Life Prediction," in *Information Science and Management Engineering (ISME), 2010 International Conference of*, 2010, pp. 369-372.

[129] D. W. Hosmer and S. Lemeshow, *Applied logistic regression* vol. 354: Wiley-Interscience, 2000.

[130] H. Liao, W. Zhao, and H. Guo, "Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model," 2006, pp. 127-132.

[131] N. Cristianini and J. Shawe-Taylor, *An introduction to support Vector Machines: and other kernel-based learning methods*: Cambridge Univ Pr, 2000.

[132] V. N. Vapnik, *The nature of statistical learning theory*: Springer-Verlag New York Inc, 2000.

[133] S. Abe, *Support vector machines for pattern classification*: Springer-Verlag New York Inc, 2010.

[134] V. N. Vapnik, "An overview of statistical learning theory," *Neural Networks, IEEE Transactions on,* vol. 10, pp. 988-999, 1999.

[135] L. Wang, *Support Vector Machines: theory and applications* vol. 177: Springer Verlag, 2005.

[136] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery,* vol. 2, pp. 121-167, 1998.

[137] A. Widodo and B. S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing,* vol. 21, pp. 2560-2574, 2007.

[138] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on,* vol. 13, pp. 415-425, 2002.

[139] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST),* vol. 2, p. 27, 2011.

[140] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, *et al.*, "WEKA---Experiences with a Java Open-Source Project," *The Journal of Machine Learning Research,* vol. 11, pp. 2533-2541, 2010.

[141] B. S. Yang and A. Widodo, "Support vector machine for machine fault diagnosis and prognosis," *Journal of System Design and Dynamics,* vol. 2, pp. 12-23, 2008.

[142] H. E. Kim, A. C. C. Tan, J. Mathew, E. Y. H. Kim, and B. K. Choi, "Integrated Diagnosis and Prognosis Model for High Pressure LNG Pump," 2009.

[143] B. Saha and K. Goebel, "Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques," 2008, pp. 1-8.

[144] W. Wang, "A prognosis model for wear prediction based on oil-based monitoring," *Journal of the Operational Research Society,* vol. 58, pp. 887-893, 2007.

[145] A. Bennane and S. Yacout, "LAD-CBM; new data processing tool for diagnosis and prognosis in condition-based maintenance," *Journal of Intelligent Manufacturing,* vol. 23, pp. 265-275, 2012.

[146] A. Ghasemi, S. Esmaeili, and S. Yacout, "Equipment's Prognostics Using Logical Analysis of Data," in *Advances in Production Management Systems. Competitive Manufacturing for Innovative Products and Services,* ed: Springer, 2013, pp. 240-247.

[147] Y. Crama, P. L. Hammer, and T. Ibaraki, "Cause-effect relationships and partially defined Boolean functions," *Annals of Operations Research,* vol. 16, pp. 299-325, 1988.

[148] H. Liu, F. Hussain, C. L. Tan, and M. Dash, "Discretization: An enabling technique," *Data mining and knowledge discovery,* vol. 6, pp. 393-423, 2002.

[149] E. Boros, P. L. Hammer, T. Ibaraki, and A. Kogan, "Logical analysis of numerical data," *Mathematical Programming,* vol. 79, pp. 163-190, 1997.

[150] H. S. Ryoo and I. Y. Jang, "Milp approach to pattern generation in logical analysis of data," *Discrete Applied Mathematics,* vol. 157, pp. 749-761, 2009.

[151] T. Bonates, P. L. Hammer, and A. Kogan, "Maximum patterns in datasets," *Discrete Applied Mathematics*, vol. 156, pp. 846-861, 2008.

[152] P. L. Hammer, A. Kogan, B. Simeone, and S. Szedmák, "Pareto-optimal patterns in logical analysis of data," *Discrete Applied Mathematics,* vol. 144, pp. 79-102, 2004.

[153] P. L. Hammer and T. O. Bonates, "Logical analysis of data—an overview: from combinatorial optimization to medical applications," *Annals of Operations Research,* vol. 148, pp. 203-225, 2006.

[154] C. Guo and H. S. Ryoo, "Compact MILP models for optimal and Pareto-optimal LAD patterns," *Discrete Applied Mathematics,* 2012.

[155] J. Han, N. Kim, B. J. Yum, and M. K. Jeong, "Pattern selection approaches for the logical analysis of data considering the outliers and the coverage of a pattern," *Expert Systems with Applications*, vol. 38, pp. 13857-13862, 2011.

[156] V. Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of operations research*, vol. 4, pp. 233-235, 1979.

[157] E. Mayoraz and M. Moreira, "On the decomposition of polychotomies into dichotomies," 1997, pp. 219-226.

[158] L. M. Moreira, "The use of Boolean concepts in general classification contexts," Universidade do Minho, Portugal, 2000.

[159] J. F. A. Herrera and M. M. Subasi, "Logical Analysis of Multi-Class Data," 2013.

[160] C. Dupuis, M. Gamache, and J.-F. Pagé, "Logical analysis of data for estimating passenger show rates at Air Canada," *Journal of Air Transport Management,* vol. 18, pp. 78-81, 2012.

[161] P. L. Hammer, A. Kogan, and M. A. Lejeune, "A logical analysis of banks' financial strength ratings," *Expert Systems with Applications,* vol. 39, pp. 7808-7821, 2012.

[162] S. Yacout, D. Salamanca, and M.-A. Mortada, "Tool and method for fault detection of devices by condition based maintenance," ed: Google Patents, 2011.

[163] M. Pintilie, *Competing risks: a practical perspective* vol. 58: John Wiley & Sons, 2006.

[164] D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modeling of Time to Event Data*: Wiley, 2011.

[165] L. D. Fisher and D. Y. Lin, "Time-dependent covariates in the Cox proportional-hazards regression model," *Annual review of public health*, vol. 20, pp. 145-157, 1999.

[166] C. A. Bellera, G. MacGrogan, M. Debled, C. T. de Lara, V. Brouste, and S. Mathoulin-Pélissier, "Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer," *BMC medical research methodology*, vol. 10, p. 20, 2010.

[167] R. Fourer, D. Gay, and B. W. Kernighan, "The AMPL book," ed: Duxbury Press, Pacific Grove, 2002.

# APPENDIX A – DISTANCE EVALUATION TECHNIQUE

Figure A-1 shows how the observations (data points) from two classes; Class 1 and Class 2, are represented in the feature space [20]. The feature $Z_2$ in the right of the figure is more efficient to separate the observations of the two classes than the feature $Z_1$ in the left. In the left of the figure, the two probability distributions overlap, because of their relatively larger standard deviation values ($\sigma_1$ and $\sigma_2$) and smaller distance between the mean values ($\mu_1$ and $\mu_2$), compared with the feature in the right of the figure. Consequently, the feature in the right has a higher discriminative power than the other one.

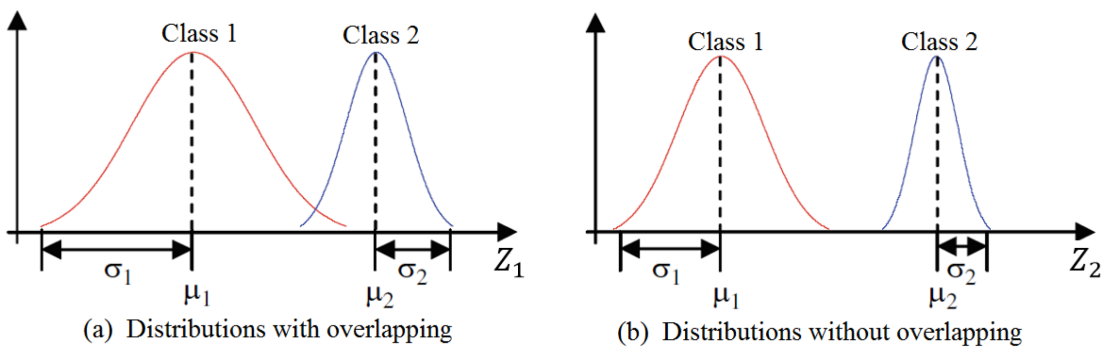


Figure A-1: Representation of two distributions using different feature spaces

The distance evaluation technique (DET) is used in this thesis as an efficient feature selection algorithm. The idea is to evaluate the discriminative power of each individual feature within a set of classes. Features that have low discriminative power are excluded from the set of the extracted features, because they may contain nonsignificant information. The features with smaller distances within the classes and larger distance among them are superior to the other features.

The DET algorithm presented in [77] is discussed as follows:

Suppose that we have a training dataset with $J$ extracted features, $C$ classes, and $M_c$ observations in each class $c$. The value of each observation with the identity $m$ at the $j^{th}$ feature in class $c$ is denoted by $q_{m,c,j}$, where $m = 1,2, \ldots . M_c$, $c = 1,2, \ldots . , C$, and $j = 1,2, \ldots . , J$.

The importance of the $j^{th}$ feature in the training dataset is assessed through the following steps.

Step 1: Calculating the average distance between the observations of the same class $c$ as:

$$d_{c,j} = \frac{1}{M_c \times (M_c - 1)} \sum_{m=1}^{M_c} \sum_{l=1}^{M_c} \left| q_{m,c,j} - q_{l,c,j} \right| \tag{A.1}$$

where $l = 1,2, \ldots . M_c, m = 1,2, \ldots . M_c$, and $l \neq m$

Step 2: Calculating the average distance for all the $C$ classes as follows:

$$d_j^W = \frac{1}{C} \sum_{c=1}^{C} d_{c,j} \tag{A.2}$$

Step 3: Calculating the variance factor of $d_j^W$ as follows:

$$v_j^W = \frac{\max(d_{c,j})}{min(d_{c,j})} \tag{A.3}$$

Step 4: Calculating the average value of the feature for all observations in the same class as follows:

$$u_{c,j} = \frac{1}{M_c} \sum_{m=1}^{M_c} q_{m,c,j} \tag{A.4}$$

Step 5: Obtaining the average distance between the distinct classes as follows:

$$d_j^b = \frac{1}{C \times (C-1)} \sum_{e=1}^{C} \sum_{f=1}^{C} \left| u_{e,j} - u_{f,j} \right| \tag{A.5}$$

where $e = 1,2, \ldots . C, \;\; f = 1,2, \ldots . C$, and $e \neq f$

Step 6: Calculating the variance factor of $u_{c,j}$ as:

$$v_j^b = \frac{\max(\left| u_{e,j} - u_{f,j} \right|)}{min(\left| u_{e,j} - u_{f,j} \right|)} \tag{A.6}$$

Step 7: Calculating the compensation factor as follows:

$$\lambda_j = \frac{1}{\dfrac{v_j^W}{\max(v_j^W)} + \dfrac{v_j^b}{\max(v_j^b)}} \tag{A.7}$$

Step 8: Calculating the ratio for the $j^{th}$ feature :

$$\alpha_j = \lambda_j \frac{d_j^b}{d_j^W} \tag{A.8}$$

Step 9: Normalizing the ratio $\alpha_j$ to get the distance evaluation factor (weight) as follows:

$$\alpha_{j-norm} = \frac{\alpha_j}{max(\alpha_j)} \tag{A.9}$$

The above steps are repeated until each feature in the dataset is assigned a weight. The features with the largest weights are selected to properly separate the $C$ classes, while the others are discarded.

# APPENDIX B –NON PARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION FOR KAPLAN-MEIER ESTIMATOR

Let $t_1 < t_1 < ... < t_r$ be the failure times (ordered chronologically) that are observed from a set of historical systems. The number of failures at the time $t_j$ is denoted by $d_j$, and $m_j$ is the the number of systems that were right-censored in the interval $[t_j, t_{j+1})$ where $t_j \leq t_r$.

Under the assumption of independence between the censoring times and the failure times, the likelihood for an observed observation at time $t_j$, presented in [163], is given as:

$$L = \prod_{j=1}^{r} \left[f(t_j)\right]^{d_j} \left[S(t_j)\right]^{m_j} \tag{B.1}$$

The hazard function at the time $t_j$ is expressed as:

$$h(t_j) = \Pr\left[T = t_j \middle| T > t_{j-1}\right] = \frac{f(t_j)}{S(t_{j-1})} \tag{B.2}$$

where $f(t_j) = \Pr[T = t_j] = F(t_j) - F(t_{j-1}) = S(t_{j-1}) - S(t_j)$

Thus, the hazard function can be written as:

$$h(t_j) = \frac{S(t_{j-1}) - S(t_j)}{S(t_{j-1})} = 1 - \frac{S(t_j)}{S(t_{j-1})} \tag{B.3}$$

Accordingly, the survival function in terms of $h(t_j)$ is represented as:

$$S(t_j) = \prod_{j=1}^{r} 1 - h(t_j) \tag{B.4}$$

According to (B.1) and (B.2), the likelihood function is expressed in terms of $h(t_j)$ and $S(t_j)$ as:

$$L(h, S; d_1, ..d_r, m_1, ..m_r) = \prod_{j=1}^{r} \left[h(t_j)\right]^{d_j} \left[S(t_{j-1})\right]^{d_j} \left[S(t_j)\right]^{m_j} \tag{B.5}$$

Based on (B.4), the likelihood function in (B.5) can be expressed in terms of $h(t_j)$ as follows:

$$\begin{aligned} L(h; d_1, ..d_r, m_1, ..m_r) &= \prod_{j=1}^{r} \left\{ \left[h(t_j)\right]^{d_j} \left[\prod_{i=1}^{j-1} 1 - h(t_i)\right]^{d_j} \left[\prod_{i=1}^{j} 1 - h(t_i)\right]^{m_j} \right\} \\ &= \prod_{j=1}^{r} \left\{ \frac{\left[h(t_j)\right]^{d_j}}{\left[1 - h(t_j)\right]^{d_j}} \left[\prod_{i=1}^{j} 1 - h(t_i)\right]^{d_j} \left[\prod_{i=1}^{j} 1 - h(t_i)\right]^{m_j} \right\} \end{aligned}$$

$$= \prod_{j=1}^{r} \left\{ \frac{\left[h(t_j)\right]^{d_j}}{\left[1 - h(t_j)\right]^{d_j}} \left[ \prod_{i=1}^{j} 1 - h(t_i) \right]^{d_j + m_j} \right\}$$

$$= \left[ \prod_{j=1}^{r} \frac{\left[h(t_j)\right]^{d_j}}{\left[1 - h(t_j)\right]^{d_j}} \right] \left[ \prod_{j=1}^{r} \prod_{i=1}^{j} [1 - h(t_i)]^{d_j + m_j} \right]$$

$$= \left[ \prod_{j=1}^{r} \frac{\left[h(t_j)\right]^{d_j}}{\left[1 - h(t_j)\right]^{d_j}} \right] \left[ \prod_{j=1}^{r} \left[1 - h(t_j)\right]^{\sum_{i=j}^{r} (d_i + m_i)} \right]$$

Let $n_j$ denotes the number of systems at risk just prior to time $t_j$, given as:

$$n_j = (d_j + m_j) + \cdots (d_r + m_r) = \sum_{i=j}^{r} (d_i + m_i)$$

It follows that:

$$L(h; d_1, .. d_r, n_1, .. n_r) = \left[ \prod_{j=1}^{r} \frac{\left[h(t_j)\right]^{d_j}}{\left[1 - h(t_j)\right]^{d_j}} \right] \left[ \prod_{j=1}^{r} \left[1 - h(t_j)\right]^{n_j} \right]$$

$$L(h; d_1, .. d_r, n_1, .. n_r) = \prod_{j=1}^{r} \left[h(t_j)\right]^{d_j} [1 - h(t_j)]^{(n_j - d_j)} \qquad \text{(B.6)}$$

The objective is to maximize $L$ with respect to $h(t_j)$. Taking the logarithm of (B.6), yields

$$l(\hat{h}; d_1, .. d_r, n_1, .. n_r) = \log L = \sum_{j=1}^{r} \{ d_j \log \left[h(t_j)\right] + (n_j - d_j) \log \left[1 - h(t_j)\right] \} \qquad \text{(B.7)}$$

Differentiating (B.7) with respect to $h(t_j)$, gives

$$\frac{\partial l(\hat{h}; d_1, .. d_r, n_1, .. n_r)}{\partial h(t_j)} = \frac{d_j}{h(t_j)} - \frac{n_j - d_j}{1 - h(t_j)} \qquad \text{(B.8)}$$

Solving for equation (B.8) equal zero, the hazard function $h(t_j)$ is estimated and given as:

$$\hat{h}(t_j) = \frac{d_j}{n_j} \qquad \text{(B.9)}$$

Therefore the estimated KM survival function is given as:

$$\hat{S}(t) = \prod_{t_j \leq t} [1 - \frac{d_j}{n_j}] \qquad \text{(B.10)}$$

# APPENDIX C – REMAINING USEFUL LIFE CALCULATION

## 1- Continuous Form

Let $T$ denotes a random variable representing the system's time to failure (TTF). It is commonly known in the reliability analysis that the Mean Time To Failure (MTTF) is calculated as:

$$MTTF = E(T) = \int_{0}^{\infty} \tau f(\tau) d\tau = \int_{0}^{\infty} S(\tau) d\tau \tag{C.1}$$

where $\tau$ is a dummy variable.

Let $T - t_k$ is a random variable that represents the remaining useful life (RUL) of $T$ at time $t_k$. The mean remaining useful life (MRUL) is the expected value of the variable $T$ given that the system survives until the time $t_k$. It is calculated as presented in [93] as follows :

$$MRUL(t_k) = E(T - t_k | T > t_k) = \frac{\int_{t_k}^{\infty} (\tau - t_k) f(\tau) d\tau}{S(t_k)} \tag{C.2}$$

$$= \frac{\int_{t_k}^{\infty} \tau f(\tau) d\tau}{S(t_k)} - \frac{t_k \int_{t_k}^{\infty} f(\tau) d\tau}{S(t_k)}$$

Since $\int_{t_k}^{\infty} f(\tau) d\tau = S(t_k)$, it follows:

$$MRUL(t_k) = \frac{\int_{t_k}^{\infty} \tau f(\tau) d\tau}{S(t_k)} - t_k \tag{C.3}$$

The MRUL at time $t_k = 0$ is equal to the MTTF that can be calculated from equation (C.3) as follows:

$$MTTF = MRUL(0) = \frac{\int_{0}^{\infty} \tau f(\tau) d\tau}{S(0)} - 0 = \int_{0}^{\infty} \tau f(\tau) d\tau$$

where $S(0) = 1$. The calculated MTTF using (C.3) is identical to that is calculated using (C.1).

The $MRUL(t_k)$ can be also calculated as presented in [94] and given as :

$$MRUL(t_k) = E(T - t_k | T > t_k) = \frac{\int_{t_k}^{\infty} S(\tau) d\tau}{S(t_k)} \tag{C.4}$$

It can be proven that equation (C.2) and equation (C.4) are identical, in the following.
The numerator of equation (C.2) is represented as:

$$Num = \int_{t_k}^{\infty} (\tau - t_k) f(\tau) d\tau$$

It is known that $f(\tau) = -\frac{dS(\tau)}{d\tau}$, thus

$$Num = -\int_{t_k}^{\infty} (\tau - t_k)\frac{dS(\tau)}{d\tau} d\tau$$

$$= -\int_{t_k}^{\infty} (\tau - t_k) dS(\tau) \qquad (C.5)$$

Let us denote $u = \tau - t_k$ and $dv = dS(\tau)$. Therefore $du = d\tau$ and $v = S(\tau)$.

After substituting $u$ and $dv$ in equation (C.5), we get

$$Num = -\int_{t_k}^{\infty} udv \qquad (C.6)$$

Integrating equation (C.6) by parts, we get

$$Num = -\left\{[uv]_{t_k}^{\infty} - \int_{t_k}^{\infty} vdu\right\}$$

$$= -\left\{[(\tau - t_k)S(\tau)]_{t_k}^{\infty} - \int_{t_k}^{\infty} S(\tau)d\tau\right\}$$

$$= -[(\tau - t_k)S(\tau)]_{t_k}^{\infty} + \int_{t_k}^{\infty} S(\tau)d\tau$$

$$Num = -\lim_{t\to\infty} \tau S(\tau) + t_k S(t_k) + t_k S(\infty) - t_k S(t_k) + \int_{t_k}^{\infty} S(\tau)d\tau$$

$$Num = -\lim_{t\to\infty} \tau S(\tau) + \int_{t_k}^{\infty} S(\tau)d\tau$$

where $S(\infty) = 0$.

It is known that $\lim_{t\to\infty} \tau S(\tau) = 0$ (the rate of decreasing the reliability function is greater than the rate of increasing the time), which is true for all distributions that have MTTF. Accordingly, we get

$$Num = \int_{t_k}^{\infty} (\tau - t_k) f(\tau) d\tau = \int_{t_k}^{\infty} S(\tau)d\tau \qquad (C.7)$$

It implies that

$$MRUL(t_k) = \frac{\int_{t_k}^{\infty}(\tau - t_k)f(\tau)d\tau}{S(t_k)} = \frac{\int_{t_k}^{\infty} S(\tau)d\tau}{S(t_k)} \qquad (C.8)$$

Eventually, the $MRUL$ can be calculated either by using equation (C.2) or equation (C.4) that are identical.

## 2- Discrete Form

For the discrete random variable $T$, the probability at time $t_k$ is defined as: $P(T \geq t_k) = P(T > t_{k-1})$. Accordingly, equation (C.3) is represented in the discrete form as:

$$MRUL(t_k) = E(T - t_k | T \geq t_k) = E(T - t_k | T > t_{k-1})$$

$$= \frac{\sum_{t_j=t_k}^{\infty} t_j \left[F\left(t_j\right) - F\left(t_{j-1}\right)\right]}{S(t_{k-1})} - t_k \qquad (\text{C}.9)$$

By considering the survival probability instead of the cumulative probability of failure, we get

$$MRUL(t_k) = \frac{\sum_{t_j=t_k}^{\infty} t_j \left[S\left(t_{j-1}\right) - S\left(t_j\right)\right]}{S(t_{k-1})} - t_k \qquad (\text{C}.10)$$

where $F\left(t_j\right) - F\left(t_{j-1}\right) = 1 - S\left(t_j\right) - 1 + S\left(t_{j-1}\right) = S\left(t_{j-1}\right) - S\left(t_j\right)$.

In a similar way, equations (C.4) is represented in the discrete form as:

$$MRUL(t_k) = E(T - t_k | T > t_{k-1}) = \frac{\sum_{t_j=t_k}^{\infty} \Delta t_j S\left(t_j\right)}{S(t_{k-1})} \qquad (\text{C}.11)$$

We can also go from equation (C.10) to equation (C.11) through some mathematical manipulations in the following way:

We can rewrite equation (C.10) as:

$$MRUL(t_k) = \frac{1}{S(t_{k-1})}\{t_k[S(t_{k-1}) - S(t_k)] + t_{k+1}[S(t_k) - S(t_{k+1})] + t_{k+1}[S(t_k) - S(t_{k+1})] + \cdots\} - t_k$$

$$= \frac{1}{S(t_{k-1})}\{t_k S(t_{k-1}) + S(t_k)[t_{k+1} - t_k] + S(t_{k+1})[t_{k+2} - t_{k+1}] + \cdots\} - t_k$$

$$= \frac{t_k S(t_{k-1})}{S(t_{k-1})} + \frac{1}{S(t_{k-1})}\{S(t_k)[t_{k+1} - t_k] + S(t_{k+1})[t_{k+2} - t_{k+1}] + \cdots\} - t_k$$

$$= t_k + \frac{1}{S(t_{k-1})}\{S(t_k)[t_{k+1} - t_k] + S(t_{k+1})[t_{k+2} - t_{k+1}] + \cdots\} - t_k$$

$$= \frac{1}{S(t_{k-1})}\{S(t_k)[t_{k+1} - t_k] + S(t_{k+1})[t_{k+2} - t_{k+1}] + \cdots\}$$

$$MRUL(t_k) = \frac{\sum_{t_j=t_k}^{\infty}[t_{j+1} - t_j]S\left(t_j\right)}{S(t_{k-1})} = \frac{\sum_{t_j=t_k}^{\infty} \Delta t_j S\left(t_j\right)}{S(t_{k-1})}$$

which is identical to equation (C.11), where $\Delta t_j = t_{j+1} - t_j$.

# APPENDIX D – PROPORTIONAL HAZARDS MODEL

The PHM is made up of two parts: the first part is the baseline hazard function, while the second part is an exponential function including all the covariates (the operating conditions and condition indicators) that affect the system's time to failure. The most common example for the baseline hazard function is the Weibull hazard function which is presented in this appendix.

Given the time-independent vector of covariates $Z_t$, the Weibull hazard function is expressed as given in [123], as:

$$\begin{aligned} h(t, Z_t) &= \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} exp(\gamma Z_t) \\ &= h_0(t)\, exp(\gamma Z_t) \end{aligned} \tag{D.1}$$

where $h_0(t)$ is baseline hazard function. The survival function of the Weibull PHM found in [164] is expressed as:

$$S(t, Z_t) = exp\left(-\int_0^{\infty} h(\tau, Z_t)\, d\tau\right)$$

$$\begin{aligned} S(t, Z_t) &= exp\left(-\left(\frac{t}{\eta}\right)^{\beta} exp(\gamma Z_t)\right) \\ &= \left[exp\left(-\left(\frac{t}{\eta}\right)^{\beta}\right)\right]^{exp(\gamma Z_t)} \\ &= [S_0(t)]^{exp(\gamma Z_t)} \end{aligned} \tag{D.2}$$

where $S_0(t)$ is the baseline survival function.

Given the lifetime data and the corresponding vectors of covariates, the three parameters $(\beta, \eta, \gamma)$ of the baseline hazard function, and the vector $\gamma$ that represents the effects of the covariates, are estimated using the maximum likelihood estimation method [125].

Based on the survival function in (D.2), the MRUL at a certain time instant $t_k$ is calculated as:

$$\begin{aligned} MRUL(t_k) &= E(T - t_k | T > t_k, Z_t) \\ &= \frac{\int_{t_k}^{\infty} (\tau - t_k) f(\tau, Z_\tau) d\tau}{S(t_k, Z_t)} = \frac{\int_{t_k}^{\infty} S(\tau, Z_\tau) d\tau}{S(t_k, Z_t)} \end{aligned} \tag{D.3}$$

The MRUL calculation is dependent on the value of $exp(\gamma Z_t)$ that is affecting the reliability function. The systems that are working under severe operating conditions are expected to fail earlier (let's call them short life systems), and the systems that are working under relaxed operating conditions are expected to fail after long time (we call them long life systems).

The application of PHM in the field of CBM has three main limitations [91, 164-166]:

- It needs some statistical assumptions about the probaility distribution of the lifetime data to build the baseline function.
- In case of time-dependent covariates, an extended PHM model which assigns a function of time to each covariate, should be used. This requires multidimensional integration when estimating the hazard or reliability function.
- The covariates must satisfy the proportional hazard (PH) assumption which is difficult or even impossible to be satisfied in the case of highly correlated covariates.

# APPENDIX E – DATA BINARIZATION

## Binarization of the numerical factors

The number of binary attributes needed to replace the numerical factor $Z$ depends on the number of distinct values of $Z$. The binarization procedure starts by ranking, in ascending order, all the distinct values of $Z$ as follows:

$$u_Z^{(1)} < u_Z^{(2)} < \cdots < u_Z^{(L)}$$

where $L$ is the total number of distinct values of $Z$ and $N$ is the total number of observations, and $L \leq N$.

A cut-point $\alpha_{Z,j}$ is introduced between each pair of values that belong to different classes. The cut-point is calculated by averaging the two values as:

$$\alpha_{Z,j} = \frac{u_Z^{(i)} + u_Z^{(i+1)}}{2} \tag{E.1}$$

where $u_Z^{(i)} \in \Omega^+$ and $u_Z^{(i+1)} \in \Omega^-$, and vice versa.

A binary attribute is then formed from each cut-point. Each cut-point $\alpha_{Z,j}$ has a corresponding binary attribute $b_{\alpha_{Z,j}}$, defined as:

$$b_{\alpha_{Z,j}} = \begin{cases} 1 & \text{if } \ u_Z \geq \alpha_{Z,j} \\ 0 & \text{if } \ u_Z < \alpha_{Z,j} \end{cases} \tag{E.2}$$

As a result of this binarization process, the number of binary attributes that make up the binarized training set is equal to the number of cut-points generated for each numerical factor in the training data set.

## Binarization of the categorical factors

The categorical factors are commonly found in the databases. An example of such factors is 'the shape' which has values: triangular, round, rectangular, and so on. The binarization of a categorical factor $Z$ is performed by assigning a binary attribute $b(Z, v)$ to each value $v$, which is defined such that:

$$b(Z, v) = \begin{cases} 1 & \text{if } \ Z = v \\ 0 & \text{otherwise} \end{cases} \tag{E.3}$$

The number of binary attributes is equal to the distinct values of the categorical factor. The special case of the categorical factors is the one that has only two distinct values. In this case, these values are assigned the binary values 0 and 1.

**Binarization of the continuous ordinal factors**

In the case where the values of the factor are ordered and comparative, they are called ordinal. For example temperature can take the values, cold, warm, and hot. In case of non-numerical ordinal factors, a numerical value is assigned to each non-numerical value. For example, if the temperature takes the values cold, warm, and hot, a value of 1 (one) is assigned cold, 2 (two) to warm, and 3 (three) to hot. The binarization of ordinal covariates is performed by comparing the assigned values with certain cut points. The number of binary attributes is equal to the number of cut points.

**Numerical Example**

The binarization procedure is illustrated through the simple example shown in Table E.1. The factors in the second and the fourth columns are numerical, while those in the third and the fifth columns are categorical ones. The factor in the last column is ordinal.

Table E.1: Non-binary data

| Class | $z_1$ | $z_2$ | $z_3$ | $z_4$ | $z_5$ |
|---|---|---|---|---|---|
| Positive class $\Omega^+$ | 1 | yellow | 3.2 | yes | cold |
| | 5 | green | 2.9 | no | warm |
| | 4 | green | 2.0 | yes | warm |
| | 5 | red | 2.2 | no | hot |
| Negative class $\Omega^-$ | 4 | red | 2.0 | yes | cold |
| | 3 | yellow | 1.4 | no | warm |
| | 5 | yellow | 0.7 | no | warm |

The first numerical factor $z_1$, is ranked in ascending order as shown in Table E.2.

Table E.2: Ranking of the numerical factor $z_1$ in ascending order

| Class | $+$ | $-$ | $+\,\&\,-$ | $+\,\&\,-$ |
|---|---|---|---|---|
| Distinct ranked values | $u_{Z_1}^{(1)}$ | $u_{Z_1}^{(2)}$ | $u_{Z_1}^{(3)}$ | $u_{Z_1}^{(4)}$ |
| Numerical value | 1 | 3 | 4 | 5 |

The first cut-point $\alpha_{Z_1,1}$ is introduced between the pair of values (1, 3) that belong to the different classes $(+,-)$. It is calculated by averaging these two values as:

$$\alpha_{Z_1,1} = \frac{u_{Z_1}^{(1)} + u_{Z_1}^{(2)}}{2} = \frac{1+3}{2} = 2$$

The second cut-point $\alpha_{Z_1,2}$ is introduced between the pair of values (3, 4), and is calculated as:

$$\alpha_{Z_1,2} = \frac{u_{Z_1}^{(2)} + u_{Z_1}^{(3)}}{2} = \frac{3+4}{2} = 3.5$$

The third cut-point $\alpha_{Z_1,3}$ is introduced between the pair of values (4, 5), and is calculated as:

$$\alpha_{Z_1,3} = \frac{u_{Z_1}^{(3)} + u_{Z_1}^{(4)}}{2} = \frac{4+5}{2} = 4.5$$

Accordingly, the factor $z_1$ is transformed into the binary attributes $b_{\alpha_{z_{1},1}}$, $b_{\alpha_{z_{1},2}}$ and $b_{\alpha_{z_{1},3}}$. The cut-point $\alpha_{Z_1,1}$ has a corresponding binary attribute $b_{\alpha_{z_{1},1}}$ with defined values:

$$b_{\alpha_{z_{1},1}} = \begin{cases} 1 & \text{if } \ u_{z_1} \geq \alpha_{z_1,j} \\ 0 & \text{if } \ u_{z_1} < \alpha_{z_1,j} \end{cases} \qquad j = 1,2,3$$

The binary attributes $b_{\alpha_{z_{1},2}}$ and $b_{\alpha_{z_{1},3}}$ are defined in the same manner, according to the corresponding cut points. As an example, the first numerical value of the factor $z_1$ which is $u_{z_1} = 1$, is transformed to the binary attributes $(0,0,0)$ since $1 < \alpha_{z_1,1}$ and $1 < \alpha_{z_1,2}$ and $1 < \alpha_{z_1,3}$. The numerical value $u_{z_1} = 4$, is transformed to the binary attributes $(1,1,0)$ since $4 \geq \alpha_{z_1,1}$ and $4 \geq \alpha_{z_1,2}$ and $4 < \alpha_{z_1,3}$. For the sake of simplicity, we change the names of the attributes $b_{\alpha_{z_{1},1}}$, $b_{\alpha_{z_{1},2}}$ and $b_{\alpha_{z_{1},3}}$, to $b_1$, $b_2$ and $b_3$, respectively. Each value of $z_1$ and its binarized value, is shown as follows:

| Numerical factor | |
|---|---|
| Classes | $Z_1$ |
| Positive class $\Omega^+$ | 1 |
| | 5 |
| | 4 |
| | 5 |
| Negative class $\Omega^-$ | 4 |
| | 3 |
| | 5 |

Binarization

| Binary attributes | | | |
|---|---|---|---|
| Classes | $b_1$ | $b_2$ | $b_3$ |
| Positive class $\Omega^+$ | 0 | 0 | 0 |
| | 1 | 1 | 1 |
| | 1 | 1 | 0 |
| | 1 | 1 | 1 |
| Negative class $\Omega^-$ | 1 | 1 | 0 |
| | 1 | 0 | 0 |
| | 1 | 1 | 1 |

The categorical factor $z_2$ takes three different values (yellow, green, and red). The binary attribute $b_4(z_2, \text{yellow})$ is assigned to the value yellow such that:

$$b_4(z_2, \text{yellow}) = \begin{cases} 1 & \text{if } \ z_2 = \text{yellow} \\ 0 & \text{otherwise} \end{cases}$$

Similarly, the binary attribute $b_5(z_2, \text{green})$ and $b_6(z_2, \text{red})$ is assigned to the value green and red, respectively such that:

$$b_5(z_2, \text{green}) = \begin{cases} 1 & \text{if } \ z_2 = \text{green} \\ 0 & \text{otherwise} \end{cases}$$

$$b_6(z_2, \text{red}) = \begin{cases} 1 & \text{if } \ z_2 = \text{red} \\ 0 & \text{otherwise} \end{cases}$$

Accordingly, the factor $z_2$ is binarized and converted into the binary attributes $b_4$, $b_5$, and $b_6$ as follows:

| Categorical factor | |
|---|---|
| Classes | $Z_2$ |
| Positive class $\Omega^+$ | yellow |
| | green |
| | green |
| | red |
| Negative class $\Omega^-$ | red |
| | yellow |
| | yellow |

Binarization

| Binary attributes | | | |
|---|---|---|---|
| Classes | $b_4$ | $b_5$ | $b_6$ |
| Positive class $\Omega^+$ | 1 | 0 | 0 |
| | 0 | 1 | 0 |
| | 0 | 1 | 0 |
| | 0 | 0 | 1 |
| Negative class $\Omega^-$ | 0 | 0 | 1 |
| | 1 | 0 | 0 |
| | 1 | 0 | 0 |

The numerical factor $z_3$ is binarized and converted into the binary attributes $b_7$ and $b_8$ in the same way as the numerical factor $z_1$, as follows:

| Numerical factor | |
|---|---|
| Classes | $Z_3$ |
| Positive class $\Omega^+$ | 3.2 |
| | 2.9 |
| | 2.0 |
| | 2.2 |
| Negative class $\Omega^-$ | 2.0 |
| | 1.4 |
| | 0.7 |

Binarization

| Binary attributes | | |
|---|---|---|
| Classes | $b_7$ | $b_8$ |
| Positive class $\Omega^+$ | 1 | 1 |
| | 1 | 1 |
| | 1 | 0 |
| | 1 | 1 |
| Negative class $\Omega^-$ | 1 | 0 |
| | 0 | 0 |
| | 0 | 0 |

The categorical factor $z_4$ is binarized and converted into only one binary attribute (the attribute $b_9$) since it takes only two values (yes, no), it is shown as follows:

| Categorical covariate | |
|---|---|
| Classes | $Z_4$ |
| Positive class $\Omega^+$ | yes |
| | no |
| | yes |
| | no |
| Negative class $\Omega^-$ | yes |
| | no |
| | no |

Binarization

| Binary attributes | |
|---|---|
| Classes | $b_9$ |
| Positive class $\Omega^+$ | 1 |
| | 0 |
| | 1 |
| | 0 |
| Negative class $\Omega^-$ | 1 |
| | 0 |
| | 0 |

The ordinal factor $z_5$ takes three different values (cold, warm, and hot). Therefore, we assign three ordinal numbers 1, 2, and 3 for cold, warm, and hot, respectively. Then we follow the same binarization procedures for transforming the numerical factors into binary attributes. First we start by ranking the assigned numerical values of the ordinal factor in an ascending order as shown in Table E.3.

Table E.3: Ranking of the ordinal factor $z_5$ in ascending order

| Class | $+\,\&\,-$ | $+\,\&\,-$ | $+$ |
|---|---|---|---|
| Distinct ranked values | $u_{z_5}^{(1)}$ | $u_{z_5}^{(2)}$ | $u_{z_5}^{(3)}$ |
| Assigned numerical value | 1 | 2 | 3 |

The first cut-point $\alpha_{Z_5,1}$ is introduced between the pair of values (1, 2), it is calculated by averaging the two values as:

$$\alpha_{Z_5,1} = \frac{u_{Z_5}^{(1)} + u_{Z_5}^{(2)}}{2} = \frac{1+2}{2} = 1.5$$

The second cut-point $\alpha_{Z_5,2}$ is introduced between the pair of values (2, 3), and is calculated as:

$$\alpha_{Z_5,2} = \frac{u_{Z_5}^{(2)} + u_{Z_5}^{(3)}}{2} = \frac{2+3}{2} = 2.5$$

Accordingly, the factor $z_5$ is transformed into the binary attributes $b_{10}$ and $b_{11}$ as follows:

| Ordinal factor | |
|---|---|
| Classes | $Z_5$ |
| Positive class $\Omega^+$ | cold |
| | warm |
| | warm |
| | hot |
| Negative class $\Omega^-$ | cold |
| | warm |
| | warm |

Binarization

| Binary attributes | | |
|---|---|---|
| Classes | $b_{10}$ | $b_{11}$ |
| Positive class $\Omega^+$ | 0 | 0 |
| | 1 | 0 |
| | 1 | 0 |
| | 1 | 1 |
| Negative class $\Omega^-$ | 0 | 0 |
| | 1 | 0 |
| | 1 | 0 |

All the binary attributes are listed in Table E.4 as follows:

Table E.4: The binary data resulting from Table E.1

| Classes | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ | $b_6$ | $b_7$ | $b_8$ | $b_9$ | $b_{10}$ | $b_{11}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Positive class $\Omega^+$ | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 |
| Negative class $\Omega^-$ | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

# APPENDIX F – PATTERN GENERATION

The process of positive pattern generation is identical to that of negative pattern generation. For the sake of simplicity, we discuss the procedure found in [150], for generating the positive patterns.

Basically, the pattern generation procedure is formulated as an MILP as follows:

The decision variables of this formulation are defined first. Given a binarized training data set composed of $n$ binary attributes, each observation $i \in \Omega^+$ or $i \in \Omega^-$ is associated with the Boolean observation vector $a_i = (a_{i,1}, a_{i,2}, \dots a_{i,n}, a_{i,n+1}, a_{i,n+2}, \dots a_{i,2n})$ whose size is $2n$, such that $a_{i,j} = 1$ if the binary attribute $b_j = 1$ and $a_{i,n+j} = 1$ if $b_j = 0$. The values $a_{i,j}$ and $a_{i,n+j}$ are mutually exclusive since $b_j$ cannot be 1 and 0 at the same time for the same observation (i.e. $a_{i,n+j} = 1 - a_{i,j}$).

Each generated pattern $p$ is associated with a Boolean pattern vector $W = (w_1, w_2, \dots w_n, w_{n+1}, w_{n+2}, \dots w_{2n})$ whose size is the same as the binary observation vector $a_i$. The elements of the pattern vector $W$ are relative to the attributes such that if $w_j = 1$ then the literal $b_j$ is included in pattern $p$. Similarly, if $w_{n+j} = 1$ then literal $\bar{b}_j$ is included in pattern $p$. A pattern $p$ cannot include both the literal $b_j$ and its negation $\bar{b}_i$ at the same time; the following condition must be satisfied:

$$w_j + w_{n+j} \leq 1 \qquad \forall\, j = 1,2, \dots n \qquad (\mathrm{F.1})$$

For the generation of a positive pattern, the Boolean vector $Y = (y_1, y_2, \dots y_{|\Omega^+|})$, whose number of elements equals the number of positive observations, is presented to indicate the coverage of the positive observations. The value of the entry $y_i$ in the vector $Y$ is defined as:

$$y_i = \begin{cases} 0 & \text{if the observation } a_i \text{ is covered by pattern } p \\ 1 & \text{otherwise} \end{cases}$$

The resulting pattern must be able to cover at least one positive observation $a_i$, at the same time it is not required to cover all the positive observations in $\Omega^+$. This condition can be formulated in the form of the following constraints:

$$\sum_{j=1}^{2n} a_{i,j}\, w_j + n y_i \geq d \qquad \forall\, i \in \Omega^+ \qquad (\mathrm{F.2})$$

A positive pattern should not cover any negative observations. The following constraints must be satisfied:

$$\sum_{j=1}^{2n} a_{i,j}\, w_j \leq d-1 \qquad\qquad \forall\, i \in \Omega^- \tag{F.3}$$

The MILP formulation of the positive pattern generation procedure is given as follows:

$$\underset{W,Y,d}{minimize} \sum_{i\in\Omega^+} y_i$$

$$s.t. \begin{cases} (\mathrm{F.1}), (\mathrm{F.2}), (\mathrm{F.3}) & \\ \sum_{j=1}^{2n} w_j = d & (\mathrm{F.4}) \\ 1 \leq d \leq n & (\mathrm{F.5}) \\ W \in \{0,1\}^n & (\mathrm{F.6}) \\ Y \in \{0,1\}^{|\Omega^+|} & (\mathrm{F.7}) \end{cases}$$

Based on the optimal solution of this formulation, the resulting pattern $p$ is constructed as:

$$p \coloneqq \bigwedge_{\substack{w_j=1 \\ j\in\{1,..n\}}} b_j \bigwedge_{\substack{w_{j+n}=1 \\ j\in\{1,..n\}}} \bar{b}_j \tag{F.8}$$

**Numerical Example**

This example illustrates the pattern generation stage in LAD. Assume the given binary dataset consist of five binary attributes. The data comprise a set of four positive observations and a set of four negative observations, as shown in Table F.1. The objective is to generate the set of strong positive and negative patterns that guarantee the coverage of all positive and negative observations.

Table F.1: Eight observations (four positive and four negative)

| Observation | Class | Binary attribues | | | | |
|---|---|---|---|---|---|---|
| | | $b_1$ | $b_2$ | $b_3$ | $b_4$ | $b_5$ |
| 1 | (Positive) $\Omega^+$ | 1 | 1 | 1 | 0 | 0 |
| 2 | | 1 | 1 | 1 | 1 | 1 |
| 3 | | 1 | 0 | 1 | 1 | 0 |
| 4 | | 1 | 1 | 1 | 1 | 0 |
| 5 | (Negative) $\Omega^-$ | 1 | 0 | 1 | 0 | 0 |
| 6 | | 0 | 0 | 0 | 0 | 0 |
| 7 | | 0 | 0 | 1 | 0 | 0 |
| 8 | | 1 | 0 | 0 | 0 | 0 |

A Boolean vector $a_i = (a_{i,1}, a_{i,2}, \dots a_{i,5}, a_{i,6}, a_{i,7}, \dots a_{i,10})$ is associated with each positive or negative observation, such that $a_{i,5+j} = 1 - a_{i,j}$, as shown in Table F.2. As an example, the Boolean vector of the first positive observation is given as:

$$a_1 = (a_{1,1}, a_{1,2}, a_{1,3}, a_{1,4}\; a_{1,5}, a_{1,6}, a_{1,7}, a_{1,8}, a_{1,9}, a_{1,10}) = (1,1,1,0,0,0,0,0,1,1)$$

The values $a_{1,1}$ and $a_{1,6}$ for example are mutually exclusive i.e. $a_{1,6} = 1 - a_{1,1} = 1 - 1 = 0$.

Table F.2: The observations and their complements (first MILP procedure)

| Observation | Class | The binary attributes in Table 5 | | | | | Complements of the binary attributes | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $a_{i,1}$ | $a_{i,2}$ | $a_{i,3}$ | $a_{i,4}$ | $a_{i,5}$ | $a_{i,6}$ | $a_{i,7}$ | $a_{i,8}$ | $a_{i,9}$ | $a_{i,10}$ |
| 1 | (Positive) | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 2 | | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 4 | | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 5 | (Negative) | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 6 | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| 7 | | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 8 | | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

## First Iteration

Now, we start the procedure of generating the first positive pattern. The objective is to generate the pattern $p_1^+$ that guarantees the maximum coverage of positive observations in the set $\Omega^+$. The problem is formulated as an MILP that minimizes the number of uncovered observations. The MILP is presented in Table F.3.

Table F.3: The first MILP iteration (generation of the positive pattern $p_1^+$)

*Iteration 1*

$$Minimize\ f = y_1 + y_2 + y_3 + y_4$$

Subject to:

$$
\begin{aligned}
w_1 + w_6 &\le 1\\
w_2 + w_7 &\le 1\\
w_3 + w_8 &\le 1\\
w_4 + w_9 &\le 1\\
w_5 + w_{10} &\le 1\\
w_1 + w_2 + w_3 + w_9 + w_{10} + 5y_1 &\ge d\\
w_1 + w_2 + w_3 + w_4 + w_5 + w_9 + 5y_2 &\ge d\\
w_1 + w_3 + w_4 + w_7 + w_{10} + 5y_3 &\ge d\\
w_1 + w_2 + w_3 + w_4 + w_{10} + 5y_4 &\ge d\\
w_1 + w_3 + w_7 + w_9 + w_{10} &\le d - 1\\
w_6 + w_7 + w_8 + w_9 + w_{10} &\le d - 1\\
w_3 + w_6 + w_7 + w_9 + w_{10} &\le d - 1\\
w_1 + w_7 + w_8 + w_9 + w_{10} &\le d - 1\\
w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9 + w_{10} &= d
\end{aligned}
$$

$$1 \le d \le 5$$

$$w_j \in \{0,1\} \quad j = 1,2,\dots 10$$

$$y_i \in \{0,1\} \quad i = 1,2,3,4$$

*Solution:*

$f = 1$

$d = 1$

$y_1 = 0, y_2 = 0, y_3 = 1, y_4 = 0$

$w_1 = 0, w_2 = 1, w_3 = 0, w_4 = 0, w_5 = 0, w_6 = 0, w_7 = 0, w_8 = 0, w_9 = 0, w_{10} = 0$

This MILP is solved by using AMPL; A Modeling Language for Mathematical Programming [167]. AMPL is not a solver, it is a language for modeling the mathematical programs. The previous

MILP is coded in AMPL and then AMPL translates the code to a solver like CPLEX or MINOS. In this example, we used CPLEX to solve this MILP and the solution is also presented in Table F.3.

The optimal solution for this MILP gives one positive pattern of degree one ($d = 1$) that covers three positive observations. The optimal value of the vector $Y = (y_1, y_2, y_3, y_4) = (0, 0,1,0)$ indicates that there is only one observation that is not covered by the pattern $p_1^+$, and that the pattern covers the first, the second, and the fourth positive observations (i.e. $cov(p_1^+) = \{1,2,4\}$). The generated pattern is constructed from this solution using equation (F.8) as $p_1^+ \coloneqq b_2$.

**Second Iteration**

All the observations in Table F.1 must be covered at least once. Accordingly, we remove the observations that are already covered by the generated patterns. In this case, the first, second, and fourth positive observations are removed. Only the third positive observation is considered in order to generate a pattern that covers that observation. The objective is to generate the second positive pattern $p_2^+$ that guarantees the coverage of the third observation (which is not covered in the first iteration), while guaranteeing the maximum coverage of positive observations in the set $\Omega^+$. Hence, the data used to generate this pattern are shown in Table F.4.

Table F.4: The observations and their complements (second MILP procedure)

| Observation | Class | The binary attributes in Table 5 | | | | | Complements of the binary attributes | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $a_{i,1}$ | $a_{i,2}$ | $a_{i,3}$ | $a_{i,4}$ | $a_{i,5}$ | $a_{i,6}$ | $a_{i,7}$ | $a_{i,8}$ | $a_{i,9}$ | $a_{i,10}$ |
| 3 | (Positive) | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 |
| 5 | (Negative) | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 |
| 6 | | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| 7 | | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 8 | | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |

The MILP formulation and its solution are given in Table F.5. This optimal solution of MILP gives one positive pattern of degree one ($d = 1$) that covers the third positive observation since $y_3 = 0$. The second positive pattern is constructed from this solution as $p_2^+ \coloneqq b_4$. However, this pattern also covers two of the removed observations $\{2,4\}$. Hence the coverage of this pattern is $Cov(p_2^+) = \{2,3,4\}$. All the positive observations are covered now, so the procedure for the generation of the positive patterns is terminated.

Table F.5: The second MILP iteration (generation of the positive pattern $p_2^+$)

*Iteration 2*

$$Minimize\ f = y_3$$

Subject to:

$$\begin{aligned}
w_1 + w_6 &\le 1 \\
w_2 + w_7 &\le 1 \\
w_3 + w_8 &\le 1 \\
w_4 + w_9 &\le 1 \\
w_5 + w_{10} &\le 1 \\
w_1 + w_3 + w_4 + w_7 + w_{10} + 5y_3 &\ge d \\
w_1 + w_3 + w_7 + w_9 + w_{10} &\le d-1 \\
w_6 + w_7 + w_8 + w_9 + w_{10} &\le d-1 \\
w_3 + w_6 + w_7 + w_9 + w_{10} &\le d-1 \\
w_1 + w_7 + w_8 + w_9 + w_{10} &\le d-1 \\
w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9 + w_{10} &= d
\end{aligned}$$

$$1 \le d \le 5$$

$$w_j \in \{0,1\} \quad j = 1,2,\ldots 10$$

$$y_3 \in \{0,1\}$$

*Solution:* $f = 0, d = 1, y_3 = 0$

$w_1 = 0, w_2 = 0, w_3 = 0\ , w_4 = 1, w_5 = 0, w_6 = 0, w_7 = 0, w_8 = 0, w_9 = 0, w_{10} = 0$

**Third Iteration**

In a similar way, the negative patterns are generated. The MILP for the generation of the negative patterns is formulated in Table F.6. The negative pattern is constructed from the resulting solution as $p_1^- \coloneqq \overline{b_2}\ \overline{b_4}$. It covers all the negative observations (i.e. $Cov(p_1^-) = \{5,6,7,8\}$). Table F.7 lists the generated positive and negative patterns.

Table F.6: The third MILP iteration (generation of the negative pattern $p_1^-$)

*Iteration 3*

$$Minimize\ f = y_5 + y_6 + y_7 + y_8$$

Subject to:

$$\begin{aligned}
w_1 + w_6 &\le 1 \\
w_2 + w_7 &\le 1 \\
w_3 + w_8 &\le 1 \\
w_4 + w_9 &\le 1 \\
w_5 + w_{10} &\le 1 \\
w_1 + w_3 + w_7 + w_9 + w_{10} + 5y_5 &\ge d \\
w_6 + w_7 + w_8 + w_9 + w_{10} \quad 5y_6 &\ge d \\
w_3 + w_6 + w_7 + w_9 + w_{10} + 5y_7 &\ge d \\
w_1 + w_7 + w_8 + w_9 + w_{10} + 5y_8 &\ge d \\
w_1 + w_2 + w_3 + w_9 + w_{10} &\le d-1 \\
w_1 + w_2 + w_3 + w_4 + w_5 + w_9 + &\le d-1 \\
w_1 + w_3 + w_4 + w_7 + w_{10} &\le d-1 \\
w_1 + w_2 + w_3 + w_4 + w_{10} &\le d-1 \\
w_1 + w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 + w_9 + w_{10} &= d
\end{aligned}$$

$$1 \le d \le 5$$

$$w_j \in \{0,1\} \quad j = 1,2,\ldots 10$$

$$y_i \in \{0,1\} \quad i = 5,6,7,8$$

*Solution:* $f = 0, d = 2, y_5 = 0, y_6 = 0, y_7 = 0, y_8 = 0$

$w_1 = 0, w_2 = 0, w_3 = 0\ , w_4 = 0, w_5 = 0, w_6 = 0, w_7 = 1, w_8 = 0, w_9 = 1, w_{10} = 0$

Table F.7: The generated positive and negative patterns

| The generated *positive* patterns | Interpretation | Covered observations *Cov(p)* |
|---|---|---|
| $p_1^+$ | $b_2$ | $1, 2, 4$ |
| $p_2^+$ | $b_4$ | $2, 3, 4$ |
| The generated *negative* patterns | Interpretation | Covered observations *Cov(p)* |
| $p_1^-$ | $\overline{b_2}\ \overline{b_4}$ | $5, 6, 7, 8$ |

The weights for the generated patterns are give as:

$$W_1^+ = \frac{cov(p_1^+)}{\sum_{i=1}^{2} cov(p_i^+)} = \frac{cov(p_1^+)}{cov(p_1^+) + cov(p_2^+)} = \frac{3}{3+3} = 0.5$$

$$W_2^+ = \frac{cov(p_2^+)}{\sum_{i=1}^{2} cov(p_i^+)} = \frac{cov(p_2^+)}{cov(p_1^+) + cov(p_2^+)} = \frac{3}{3+3} = 0.5$$

$$W_1^- = \frac{cov(p_1^-)}{cov(p_1^-)} = \frac{4}{4} = 1.0$$